Acknowledgments and Disclosures. This research received no external financial support. The authors have no relevant financial relationships or other potential conflicts of interest related to the research described in this paper.

Subjective Expectations for Variance and Skewness: Evidence from Analyst Reports *

Shuaiyu Chen† Shuaiqi Li‡ Yucheng Yang§ Yaping Zheng¶

August 16th, 2026

Abstract

How do investors form their beliefs about the higher moments of individual stock returns? We construct *subjective* expectations for variance and skewness from analysts' scenario-based price targets over 180,000 reports from ten major brokerages. These expectations predict future realized moments controlling for option-implied and past realized moments and generate profitable option strategies. Hedge funds increase stock option positions for which variance expectations are higher, consistent with sophisticated investors acting on this information. Using within-report analyst beliefs about the three return moments, we find a *positive* subjective variance–return relation and a *negative* subjective skewness–return relation. Firm characteristics explain 34% and 7% of variations in variance and skewness expectations, respectively. The explanatory power increases to 54% and 29% after incorporating scenario-specific narratives identified by large language models and machine learning methods, with topics driving variance versus skewness being distinct.

Keywords: Subjective expectations, Variance, Skewness, Narratives, Large language models

*We are grateful for the comments from Justin J. Case (discussant), Ing-Haw Cheng, Ricardo Delao, Steve Heston (discussant), Kewei Hou, Gang Li, Cynthia Yin, and conference participants at the CBOE Volatility and Derivative Conference and New Zealand Finance Meeting. All errors are our own.

†Darden School of Business, University of Virginia, ChenS@darden.virginia.edu

‡Department of Economics and Finance, City University of Hong Kong, shuaiqli@cityu.edu.hk

§School of Accountancy, Chinese University of Hong Kong, yuchengyang@cuhk.edu.hk

¶Alberta School of Business, University of Alberta, yaping2@ualberta.ca

1. Introduction

How do market participants form beliefs about the distribution of future stock returns? While a growing literature has made progress on subjective expectations for *the mean* of returns (see e.g., Brav and Lehavy, 2003; Ben-David, Graham, and Harvey, 2013; Nagel and Xu, 2022; Giglio, Maggiori, Stroebel, and Utkus, 2021), subjective expectations for *higher return moments* remain largely a black box. This gap matters: variance and skewness are central to portfolio allocation and asset pricing (see e.g., Merton, 1987; Harvey and Siddique, 2000; Ang, Hodrick, Xing, and Zhang, 2006; Barberis and Huang, 2008), yet direct measures of investors' subjective beliefs about these moments are not available at the firm level.[1] Researchers instead often rely on option-implied moments as proxies, which embed risk premia that can drive a wedge between these measures and investors' true beliefs.[2] Beyond these measurement challenges, an even more fundamental question is *how* investors form their higher-moment beliefs: are their expectations informative about future realized moments, and if so, what types of information do they rely on? Narratives have recently been recognized as central to understanding investor expectations (see e.g., Manela and Moreira, 2017; Shiller, 2017; Hirshleifer, 2020; Goetzmann, Kim, and Shiller, 2022). However, virtually no textual data connect an individual's narrative reasoning to that same individual's beliefs about the *shape* of the return distribution. Even where such textual content exists, studying it at scale has been infeasible until the recent development of large language models (LLMs) (see Campbell and Ramadorai, 2026).

We address these challenges using a uniquely suited data source: scenario-based equity research reports. A growing number of brokerage firms require their analysts to provide not only a base-case price target but also bull-case and bear-case targets, along with narrative discussions of the firm-specific, industry, and macroeconomic factors that drive each scenario. Each report therefore bundles three distinct types of information in a single document: quan-

[1] Existing work estimates investor beliefs about market-level variance from the dispersion of survey-based market return forecasts (Ben-David, Graham, and Harvey, 2013; Lochstoer and Muir, 2022), but firm-level measures are absent.

[2] On the variance risk premium, see Bollerslev, Tauchen, and Zhou (2009) and Carr and Wu (2009). On the skewness risk premium, see Bakshi, Kapadia, and Madan (2003), Kozhan, Neuberger, and Schneider (2013), and Orlowski, Schneider, and Trojani (2023).

titative scenario price targets, from which we construct subjective variance and skewness; the analyst's expected return, derived from the base-case target relative to the current price; and the textual reasoning behind each scenario. We collect reports from the largest ten brokerage firms that mandate or regularly employ scenario-based analysis. Our sample comprises over 180,000 reports spanning 2007 to 2024, covering a broad cross-section of U.S. equities across market capitalizations and sectors.[3] The multi-broker structure also allows us to construct consensus measures and to examine cross-broker heterogeneity. To the best of our knowledge, no other data source jointly provides analyst-identified beliefs about the first three moments of a stock's return distribution together with the narrative reasoning underlying those beliefs. This unique combination makes it possible, for the first time, to measure subjective higher-moment expectations and to understand how they are formed.

We estimate the analyst's expectation of variance and skewness for the stock by calibrating a binomial tree model to each analyst's scenario forecasts over the following 12 months. Because analysts set price targets to reflect their assessment of expected outcomes (not risk-neutral pricing), the resulting expectations are free from the risk premia that contaminate option-implied moments. We find that these expectations are highly informative of future realized moments and can be used to form profitable option strategies betting on the corresponding higher moments.

To study what information analysts weigh under bear and bull scenarios, we employ LLMs, combined with human validation, to classify scenario text into 32 sub-topics spanning firm-level (e.g., revenue/earnings, risk management), industry-level (e.g., regulation/policy, market share), and macroeconomic (e.g., GDP, monetary policy, geoeconomic risk) categories, and compute metrics that capture the share of narrative attention devoted to each topic. Using state-of-the-art machine-learning approaches, we then identify the key topics driving higher-moment

[3] A few prior studies using scenario-based analyst forecasts have relied exclusively on reports from Morgan Stanley (e.g., Joos, Piotroski, and Srinivasan, 2016), which pioneered this framework in 2007 in response to competitive pressures and regulatory scrutiny of Wall Street research. Weyns, Perez, Hurewitz, and Jenkins (2011) describe the procedures Morgan Stanley used to institutionalize the framework, including tying its adoption to analysts' compensation. Other brokerage firms such as Barclays, RBC, Jefferies, and UBS subsequently adopted similar approaches.

beliefs and find that a variety of narratives, especially those related to bear-scenario information, offer significant explanatory power for analysts' higher moment expectations, substantially exceeding that of option-implied moments, realized moments, and common firm characteristics alone.[4]

We establish three sets of results. The first concerns the informativeness of analyst higher moment expectations. In panel regressions at the report level, the analyst variance expectation (IV^{Analyst}) significantly predicts next-year realized variance even after controlling for option-implied variance, past realized variance, firm characteristics, and broker and year-month fixed effects. The results for the analyst skewness expectation ($Skew^{\text{Analyst}}$) are more striking: it is the only significant predictor of future realized skewness, whereas option-implied skewness and past realized skewness carry no predictive power, consistent with the well-known difficulty of forecasting idiosyncratic skewness (see e.g., Langlois, 2020; Chen and Li, 2025). We also find that the predictive power of analyst expectations is pervasive across brokers, so our results are not driven by any single broker. This informativeness extends to option markets. IV^{Analyst} positively predicts cross-sectional straddle returns and $Skew^{\text{Analyst}}$ positively predicts returns on skewness assets (Bali and Murray, 2013), even after controlling for a comprehensive set of option return predictors. A long-short straddle strategy sorted by consensus IV^{Analyst} (orthogonalized from other option predictors) generates a monthly return of 4.71% ($t = 6.11$) with an annualized Sharpe ratio of 1.25; a skewness-asset strategy sorted by consensus $Skew^{\text{Analyst}}$ yields 0.37% per month ($t = 3.24$, Sharpe ratio 0.62). These returns are robust across equal- and value-weighted portfolios and within the S&P 500 and Russell 1000 universes.

We further exploit the structure of the price target to decompose the analyst's subjective variance and skewness into a valuation component and an earnings component. Because many analysts derive their price targets as the product of a valuation multiple and an EPS forecast in their scenario analysis, we extract scenario-specific multiples and EPS forecasts to isolate the contribution of each source to the informativeness of analyst higher-moment expectations.

[4] A few contemporaneous studies also employ LLMs to examine analyst reports (see e.g., Bastianello, Decaire, and Guenzel, 2024, 2025; Ke, 2024).

The valuation multiple component significantly predicts future realized variance, whereas the EPS component does not, suggesting that the predictive power for variance is primarily driven by analysts' superior ability to assess uncertainty about how the market will price the firm rather than about its earnings trajectory. For skewness, both components are independently significant, indicating that valuation and fundamental channels each contribute to analysts' assessments of the asymmetry in the return distribution.

We provide further evidence that these analyst expectations carry real information content by examining whether sophisticated investors may trade on them. Using institutional option ownership data from SEC 13F filings, we focus on hedge funds and measure their call, put, and straddle positions as a percentage of shares outstanding (see, Aragon, Chen, and Shi, 2022; Chen and Li, 2025). In panel regressions with firm and year-month fixed effects, the analyst variance expectation is positively and significantly associated with hedge fund call, put, and straddle ownership, even after controlling for option-implied variance, past realized variance, and firm characteristics. The significance of straddle ownership, defined as the minimum of call and put ownership, indicates that these are paired positions designed to isolate volatility exposure rather than directional bets. These results suggest that hedge funds actively incorporate the volatility information embedded in analyst scenario forecasts into their derivative trading strategies.

The second set of results exploits the joint structure of our data to test classical asset pricing theories. Merton (1987) predicts that expected returns should increase with expected variance when investors hold underdiversified portfolios, yet Ang, Hodrick, Xing, and Zhang (2006) document a puzzling negative risk-return relationship using realized return data. A key challenge is that realized returns and realized variance are noisy proxies for their ex-ante counterparts and may diverge from subjective expectations (see e.g., Nagel and Xu, 2023; Jensen, 2023). Because each analyst report simultaneously provides beliefs about the first, second, and third moments, we can test the risk–return and skewness–return trade-offs using a single analyst's beliefs, avoiding the mismatch between subjective expectations and realized outcomes that plagues conventional tests. We find a significant *positive* risk–return trade-off that sur-

vives controls for firm characteristics and holds within each of the largest brokerage firms. For skewness, theories predict that investors accept lower expected returns for positively skewed assets (Brunnermeier, Gollier, and Parker, 2007; Barberis and Huang, 2008; Mitton and Vorkink, 2007); we document a strong *negative* skewness–return relationship, consistent with these theories. When both return and risk expectations come from the same source, our data recover the trade-offs that classical theories predict but that realized data often reject.

The third set of results investigates how analysts form their higher-moment expectations by examining the narrative content of their scenario discussions. Standard observable characteristics (market-based moments and firm characteristics) leave most of the variation in analyst expectations unexplained, with adjusted R^2 of only 33.7% for variance and 6.7% for skewness, suggesting that qualitative, firm-specific, industry, and macro assessments may play a central role in scenario-based equity research. Our LLM-based textual analyses show that analysts devote the most narrative attention to revenue/earnings, valuation multiples, and risk management among firm-level topics, with regulation/policy and market share leading the industry category and GDP/consumer spending and macro uncertainty dominating the macro category.

To identify which topics genuinely drive analysts' higher-moment beliefs, as opposed to merely receiving the most attention, we employ a machine-learning model, Boosted Regression Trees combined with SHAP (SHapley Additive exPlanations) values.[5] Incorporating narrative topics dramatically improves explanatory power of analysts' higher-moment expectations. The full model, combining implied and realized moments with all 64 topic features (32 sub-topics $\times$ 2 scenarios), achieves an out-of-sample R^2 of 54% for variance and 29% for skewness, substantially exceeding the explanatory power of market moments and firm characteristics alone. This large improvement confirms that the modest explanatory power of the determinants regressions reflects the absence of qualitative narrative inputs rather than noise in analyst expectations.

Notably, the SHAP-based importance rankings diverge from these narrative attention rankings, indicating that topics receiving modest attention space in the narrative discussion can

[5]These methods have recently been adopted in machine-learning research in finance (see, e.g., Gu, Kelly, and Xiu, 2020; Lundberg and Lee, 2017; Aragon and Chen, 2024).

nonetheless exert outsized influence on expectations if they reveal analysts' unobservable opinions. The SHAP analysis also shows that the topics most important for variance expectations differ from those driving skewness. Valuation multiples and profit margins are common drivers of both, but intangible capital and revenue/earnings emerge as distinctively important for variance, whereas credit lines rank notably higher for skewness. At the industry and macro level, market share and COVID-19 are substantially more influential for skewness than for variance. This asymmetry in the information sets driving variance versus skewness reinforces the importance of examining investors' beliefs about the entire return distribution since they reflect distinct dimensions of analysts' qualitative assessments.

Our paper contributes to several strands of the literature. First, we contribute to research on subjective investor beliefs by developing novel firm-level measures of subjective variance and skewness expectations that are free from risk premia. Whereas option-implied moments are contaminated by variance and skewness risk premia and survey-based measures are available only at the market level (Ben-David, Graham, and Harvey, 2013; Lochstoer and Muir, 2022), our scenario-based analyst measures provide direct, firm-specific beliefs about higher moments. We propose a binomial tree calibration that produces moment estimates across ten brokerages over nearly two decades, substantially broader coverage than prior single-broker studies.

Second, we demonstrate that analyst expectations of the second and third moments are informative despite the well-documented biases in first-moment forecasts (Brav and Lehavy, 2003). While subjective expected returns from price targets are optimistically biased, our subjective expected variance and skewness measures predict both future realized moments and option returns in the cross-section of firms, generating economically large and statistically significant trading profits. A possible explanation is that level biases in bull, base, and bear targets partially cancel when computing central moments. Our findings connect to recent work on the information content of analyst research (e.g., Bastianello, Decaire, and Guenzel, 2024, 2025; Ke, 2024) and to asset pricing studies exploiting option-market signals for return prediction (Cao, Han, Li, Yang, and Zhan, 2024).

Third, using the same analyst's joint expectations for return, variance, and skewness, we

provide clean tests of the mean-variance and mean-skewness trade-offs. The positive risk–return and negative skewness–return relationships we document are consistent with classical asset pricing theories (Merton, 1987; Brunnermeier, Gollier, and Parker, 2007; Barberis and Huang, 2008) and underscore the importance of distinguishing subjective expectations from realized outcomes in asset pricing tests (Nagel and Xu, 2023; Jensen, 2023). This result also adds to a growing literature that uses subjective beliefs to revisit fundamental asset pricing puzzles.

Fourth, our work relates to the literature on the noise and informativeness of analyst expectations. De Silva and Thesmar (2024) decompose analyst forecast errors into noise and bias components and show that noise is a quantitatively important feature of expectations, with implications for asset prices and trading volume. Our finding that analyst higher-moment expectations are informative despite well-known biases in point forecasts is consistent with this view: central moments computed from scenario price targets are less susceptible to level biases because optimistic shifts in bull, base, and bear targets partially cancel. Relatedly, Cao, Jiang, Wang, and Yang (2024) demonstrate that combining human analyst judgment with AI-based predictions yields stock analyses that outperform either alone, highlighting the complementary roles of human expertise and machine learning in processing financial information. Our narrative analysis, which employs LLMs to extract and classify the qualitative reasoning behind analysts' scenario assessments, provides a concrete application of this human–AI complementarity in the context of higher-moment belief formation.

Finally, we contribute to the emerging literature that leverages LLMs and machine learning for economics and finance applications. For example, Jiang, Kelly, and Xiu (2024) and Lopez-Lira and Tang (2023) use LLMs on financial news to predict stock returns; Li, Mai, Shen, Yang, and Zhang (2023) extract corporate culture from analyst reports; Jha, Qian, Weber, and Yang (2024) analyze corporate investment information from conference call transcripts; Chen, Peng, and Zhou (2024) analyze social media posts to decode individual trading strategies. Our narrative-based approach provides a practical framework that links analyst scenario discussions to beliefs about higher moments and reveals that analysts form risk perceptions by synthesizing

quantitative market signals with qualitative assessments of firm, industry, and macroeconomic conditions. The finding that narrative topics explain substantially more variation in expectations than standard characteristics, and that the information sets driving variance and skewness expectations are distinct, offers new insights into the process of expectation formation.

The remainder of this paper proceeds as follows. Section 2 describes the data and variable construction. Section 3 presents the empirical results. Section 4 analyzes the narrative content of analysts' information sets. Section 6 concludes.

2. Data and Variables

2.1. Stock and Option Data

Stock prices and returns come from CRSP, and firm-level accounting data from Compustat. Option price data are obtained from the OptionMetrics Ivy DB database, which provides daily closing bid and ask quotes for U.S. equity options.

2.2. Scenario-Based Price Targets and Narratives

A growing number of brokerage firms require their equity analysts to produce scenario-based valuations alongside standard price targets. Morgan Stanley was among the first to adopt this practice, developing its Risk-Reward Framework following the heightened market uncertainty of the early 2000s and increasing regulatory pressure on Wall Street research. As described by Joos, Piotroski, and Srinivasan (2016), the framework was initially introduced in 2003 within Morgan Stanley's European research division under then-head of research Juan-Luis Perez. After Perez's promotion to global research director in 2006, the framework was implemented firmwide in January 2007.[6] The motivation was to move beyond single-point price targets by requiring analysts to specify bull-case and bear-case price estimates that quan-

[6] Weyns, Perez, Hurewitz, and Jenkins (2011) document the institutional procedures Morgan Stanley employed to ensure consistent application, including linking adoption to analysts' performance evaluations and compensation.

tify a stock's upside and downside potential, along with narrative discussions of firm-specific, industry, and macroeconomic factors that the analyst views as critical for each scenario. Other brokerage firms subsequently adopted similar frameworks, either independently or to differentiate their equity research in a competitive environment.

Our analyst reports are sourced from the Refinitiv LSEG Workspace. We identify the ten largest brokerage firms that mandate or regularly employ scenario-based analysis in their U.S. equity coverage. In the appendix, we provide some examples for equity research reports from Barclays, Morgan Stanley, and RBC (see, Fig. A.1). For each report, we use Gemini Flash 2.0 to identify the front-page information such as the report release date, the date of current stock price, ticker, analyst names, emails, cell numbers, as well as the bull, base, and bear targets (usually over the next 12 months) with narratives from the corresponding section on scenario analysis. We verify accuracy by manually cross-checking a random subset against the original reports.

Firms covered by these brokerage houses are merged with the CRSP/Compustat universe. The combined sample spans June 2007 to December 2024 and comprises 180,858 reports with bull, base, bear price targets and narratives, covering 3,496 unique U.S. public firms. Figure 1 illustrates the sample's breadth over time: the number of firms with active analyst coverage (defined as having at least one report within the preceding 12 months) grows from about 200 in 2007 to nearly 2,000 by 2024, accounting for approximately 80% of total CRSP common-stock market capitalization since 2012. In order of sample size by broker, they are: Barclays (entering our sample in April 2011; 43,639 reports), Morgan Stanley (June 2007; 32,239), RBC (January 2008; 31,396), Jefferies (March 2011; 22,745), Credit Suisse (January 2010; 17,766), UBS (December 2010; 12,125), BMO (August 2016; 11,858), Wells Fargo (July 2014; 6,103), Evercore (August 2015; 1,614), and Bank of America (July 2011; 1,373).

2.3. Constructing Analyst Subjective Variance and Skewness

From each report we extract three price forecasts: the bull-case price (*Bull*), the base-case price target (*Target*), and the bear-case price (*Bear*). We infer the analyst's subjective expectations of return variance and skewness by fitting a simple binomial tree model. The stock price is assumed to move to the bull price with probability p or to the bear price with probability $1 - p$ over the next 12 months, where p is pinned down by:

$$p \cdot Bull + (1 - p) \cdot Bear = Target. \tag{1}$$

Importantly, because analysts' price targets represent their assessment of expected outcomes rather than risk-neutral pricing, this probability reflects the analyst's real-world beliefs. Given p, we compute the variance and skewness of $\log R_{i,t\to t+12}$, where $R_{i,t\to t+12}$ denotes firm i's gross return over the next 12 months, calculated as the scenario price (bull or bear) divided by the current stock price. We denote the resulting measures IV^{Analyst} and $Skew^{\text{Analyst}}$.

Because multiple brokerage firms may cover the same stock at different points in time, we aggregate individual analyst signals into consensus measures in the analysis of option return predictability, which operates at a monthly frequency. For each firm-month, we identify all analyst reports issued within the preceding 12 months from any contributing broker. We require at least two contributing brokers whose most recent reports are no more than six months apart.

We compute two consensus variants. The *simple consensus* is the equal-weighted average of the most recent forecast from each contributing broker. The *recency-weighted consensus* weights each broker's forecast by $1/(\text{forecast lag} + 1)$, where the forecast lag is measured in months, so that more recent signals receive greater weight while still benefiting from cross-broker diversification.

2.4. *Option Variables*

To benchmark these subjective measures of higher moments, we construct the corresponding realized and option-implied moments. Realized variance ($RV_{i,t\to t+12}$) is the sum of squared daily log returns over the following 252 trading days. Realized skewness ($RSkew_{i,t\to t+12}$) is the standardized third central moment of daily log returns over the same window. Option-implied variance (IV^{Option}) and skewness ($Skew^{\text{Option}}$) are computed from 365-day implied volatilities in the OptionMetrics Volatility Surface database, following the methodology of Bakshi, Kapadia, and Madan (2003).

We construct returns on straddle and skewness-asset portfolios using equity options. Portfolios are formed on the third Friday of each month and held to expiration (typically the third Friday of the following month). The straddle consists of one at-the-money call and one at-the-money put with the same strike, selected as the strike closest to the current stock price. Following Bali and Murray (2013), the skewness asset comprises a long position in an out-of-the-money (OTM) call, a short position in an OTM put, and a position in the underlying stock, constructed to be both delta- and vega-neutral so that it isolates exposure to the third moment. The OTM put (call) is the contract with delta closest to -0.1 (0.1). Letting Δ_P (Δ_C) and ν_P (ν_C) denote the delta and vega of the put (call), the portfolio holds one OTM call, $-\nu_C/\nu_P$ OTM puts, and $-(\Delta_C - \nu_C/\nu_P \cdot \Delta_P)$ shares of the underlying. Skewness asset returns are computed as $r^{\text{Skew}} = (\text{Payoff} - \text{Price})/|\text{Price}|$, where the absolute value accounts for the possibility that the initial price is negative.[7] We apply standard filters to ensure tradability: we exclude options with zero open interest or zero bid prices, options violating standard no-arbitrage bounds, options with missing implied volatilities or deltas, options on stocks priced below \$5, and options subject to stock splits during the holding period.

Control variables for the option return analysis largely follow Heston, Jones, Khorram, Li, and Mo (2023). They include: log market capitalization, idiosyncratic volatility computed following Cao and Han (2013) from a 22-day rolling Fama–French three-factor regression,

[7] Deltas and vegas are from OptionMetrics. Option prices are computed at the bid-ask midpoint.

the volatility deviation of Goyal and Saretto (2009) (the log difference between historical and at-the-money implied volatility), the implied volatility term spread following Vasquez (2017), the IV smirk slope (the difference between 30-day call and put implied volatilities at deltas of 0.3 and -0.3, related to the risk-neutral skewness predictor in Bali and Murray (2013)), the volatility of volatility (Horenstein, Vasquez, and Xiao, 2023), the lagged option return (short-term reversal), and option return momentum (the average return over months $t-12$ to $t-2$).

2.5. *Summary Statistics*

Table 1 presents the distributional properties of our key variables. Panel A reports statistics at the analyst report level. The mean of IV^{Analyst} is 0.187, which closely tracks realized variance (mean 0.202), with comparable percentile distributions. Option-implied variance, at 0.269 on average, substantially exceeds realized variance, a gap consistent with the variance risk premium documented in the options literature. The alignment between IV^{Analyst} and realized variance indicates that our subjective measure is not inflated by risk premia.

The pattern is similar for skewness. Realized skewness averages -0.292, while option-implied skewness is considerably more negative at -0.544, consistent with the finding in Bakshi, Kapadia, and Madan (2003) that risk-neutral skewness is more negative than realized skewness in economies with fat-tailed returns. $Skew^{\text{Analyst}}$ has a mean of -0.357, falling between the option-implied and realized values. Unlike $Skew^{\text{Option}}$, which remains negative throughout most of the distribution, $Skew^{\text{Analyst}}$ turns positive at the 95th percentile, mirroring the behavior of realized skewness. The analyst's average expected return (r^{Analyst}) is 14.3%, reflecting the well-documented optimism bias in analyst forecasts.

Panel B disaggregates by brokerage firm. All ten brokers exhibit IV^{Analyst} values closer to realized variance than to option-implied variance. There is, however, meaningful cross-broker variation. Morgan Stanley reports the highest mean IV^{Analyst} (0.268) and the least negative mean $Skew^{\text{Analyst}}$ (-0.107). Barclays and RBC report more negative skewness expectations (-0.469 and -0.525, respectively). These differences likely reflect heterogeneity in analytical

methodologies, coverage universes, and internal valuation frameworks across firms.

Panels C and D of Table 1 report summary statistics for the straddle sample (108,905 firm-months) and skewness-asset sample (103,769 firm-months), along with several well-known option return predictors.

3. Empirical Results

3.1. Predicting Realized Variance and Skewness

If analysts collect and process information about firm-specific risks, their subjective expectations should contain signals about future return variance and skewness. We test this hypothesis by estimating panel regressions at the broker–firm–date level:

$$RV_{i,t\to t+12} = \alpha + \beta\, IV^{\text{Analyst}}_{b,i,t\to t+12} + \gamma\, \text{Controls}_{i,t} + \delta_b + \mu_t + \epsilon_{b,i,t},$$

$$RSkew_{i,t\to t+12} = \alpha + \beta\, Skew^{\text{Analyst}}_{b,i,t\to t+12} + \gamma\, \text{Controls}_{i,t} + \delta_b + \mu_t + \epsilon_{b,i,t},$$

where b indexes the broker, i the firm, and t the date of the current stock price reported in the analyst report (i.e., the date at which expectations are formed). The dependent variables are annualized realized variance and realized skewness computed from daily returns over the 12 months following date t. δ_b and μ_t denote broker and year-month fixed effects, respectively; the most demanding specification replaces them with broker×year-month fixed effects $\delta_{b,t}$. Controls include the corresponding option-implied moments (IV^{Option}, $Skew^{\text{Option}}$), past realized moments ($RV_{i,t-12\to t}$, $RSkew_{i,t-12\to t}$), and firm characteristics (market beta, log market capitalization, log book-to-market, asset growth, and gross profitability-to-assets). Standard errors are three-way clustered by broker, firm, and year-month. All independent variables are winsorized at the 1st and 99th percentiles.

Table 2 reports the results. Panel A examines realized variance. In column (1), IV^{Analyst} alone predicts future realized variance with a coefficient of 0.566 ($t = 12.99$). Column (2) adds

the option-implied and past realized counterparts along with firm characteristics: the coefficient on IV^{Analyst} declines to 0.180 but remains highly significant ($t = 8.11$), while IV^{Option} (0.350, $t = 12.13$) and past RV (0.158, $t = 2.93$) also contribute significantly. Column (3) introduces separate broker and year-month fixed effects, and column (4) replaces them with broker×year-month fixed effects, absorbing all time-varying broker-level heterogeneity. Even in this saturated specification, IV^{Analyst} retains a coefficient of 0.159 ($t = 8.33$), confirming that analysts contribute variance information incremental to option markets, historical volatility, firm attributes, and broker-specific time effects.

Panel B turns to realized skewness. In column (1), $Skew^{\text{Analyst}}$ positively predicts future skewness with a coefficient of 0.089 ($t = 8.48$). Notably, option-implied skewness ($t = 0.32$) and past realized skewness ($t = 0.33$) are statistically insignificant in column (2), consistent with the well-documented difficulty of forecasting return skewness (see, e.g., Chen and Li, 2025). In column (4) with broker×year-month fixed effects, $Skew^{\text{Analyst}}$ remains the sole significant predictor (coefficient 0.075, $t = 4.12$). This result underscores the distinctive value of subjective skewness expectations extracted from analyst reports, a dimension of beliefs that existing market-based measures fail to capture.

A distinctive feature of our dataset is its coverage of the largest 10 brokerage firms distributing equity research reports with scenario analysis on Refinitiv platform, whereas prior studies of scenario-based analyst forecasts have relied exclusively on Morgan Stanley reports (e.g., Joos, Piotroski, and Srinivasan, 2016). Table 3 exploits this breadth by estimating the baseline specification separately for each broker. Panel A shows that IV^{Analyst} significantly predicts realized variance for nine of the ten brokers, with coefficients ranging from 0.075 (UBS) to 0.355 (BofA). Panel B reveals that $Skew^{\text{Analyst}}$ is significant for six brokers, despite the inherently low predictability of skewness. These broker-level results demonstrate that the informational content of scenario-based forecasts is not an artifact of any single brokerage's methodology or coverage universe, but rather a general property of how analysts form higher-moment expectations.

3.2. *Predicting Option Returns*

Section 3.1 establishes that analyst expectations predict future realized moments. We now ask whether they also predict option returns—that is, whether analyst signals contain information not yet incorporated into option prices. A straddle, consisting of an at-the-money call and put at the same strike, is a volatility-sensitive position whose payoff is linked to realized volatility relative to option-implied volatility. If IV^{Analyst} contains variance information beyond what option prices already incorporate, it should predict straddle returns. Analogously, the skewness asset of Bali and Murray (2013)—a delta- and vega-neutral portfolio of out-of-the-money options that isolates exposure to the third moment—should be predictable by $Skew^{\text{Analyst}}$ if analysts' skewness beliefs are not fully spanned by option markets. Both portfolios are formed on the third Friday of each month and held to expiration, as described in Section 2.4.

We test these predictions using monthly Fama and MacBeth (1973) cross-sectional regressions:

$$r^{\text{Straddle}}_{i,t+1} = \alpha_t + \beta_t \, IV^{\text{Analyst}}_{i,t\to t+12} + \gamma_t \, \text{Controls}_{i,t} + \epsilon_{i,t+1},$$

$$r^{\text{Skew}}_{i,t+1} = \alpha_t + \beta_t \, Skew^{\text{Analyst}}_{i,t\to t+12} + \gamma_t \, \text{Controls}_{i,t} + \epsilon_{i,t+1},$$

where the analyst signal is the firm-month consensus, constructed as either the equal-weighted or recency-weighted average across brokers covering the same firm. Controls include the option-related predictors described in Section 2.4: log market capitalization, idiosyncratic volatility, volatility deviation, IV term spread, IV smirk slope, volatility of volatility, the lagged option return (short-term reversal), and 12-2 month average option return (momentum). All independent variables are cross-sectionally standardized to zero mean and unit variance each month, so coefficients reflect the effect of a one-standard-deviation change. Standard errors use Newey–West with 12 lags.

Table 4 presents the results. Panel A shows that the equal-weighted consensus IV^{Analyst} positively predicts next-month straddle returns with a coefficient of 1.145% per month (t = 2.67). The recency-weighted consensus, which gives more weight to recently updated forecasts,

yields a comparable 1.036% ($t = 2.61$). The signal is robust across individual brokers. We focus on the three largest brokers with sufficiently large firm coverage over a long sample period (see Table 4): Morgan Stanley (1.850%, $t = 3.36$), Barclays (1.640%, $t = 4.29$), and RBC (0.675%, $t = 1.88$). Panel B shows that consensus $Skew^{\text{Analyst}}$ predicts skewness asset returns (0.127%, $t = 2.85$), with Morgan Stanley's signal being particularly strong (0.232%, $t = 2.61$). These results indicate that analyst expectations carry information about future volatility and skewness that is not fully reflected in option prices.

To gauge economic significance, we implement quintile portfolio sorts. A key feature of our sorting procedure is that we first orthogonalize the analyst signal with respect to the full set of option-market controls used in the Fama–MacBeth regressions. Specifically, each month we cross-sectionally regress the analyst signal on all controls and retain the residual. This orthogonalization step ensures that the resulting portfolio spread is not driven by mechanical correlations with well-known option return predictors such as volatility deviation, volatility of volatility, or option momentum, but instead isolates the incremental information unique to analyst expectations. We then sort option portfolios into quintiles based on the orthogonalized signal and form both equally weighted and value-weighted long-short portfolios that buy the highest quintile and sell the lowest.

Table 5 reports the portfolio results. Panel A shows that the consensus straddle long-short portfolio earns 4.71% per month ($t = 6.11$), corresponding to an annualized Sharpe Ratio (SR) of 1.25. Performance is strong for individual brokers as well: Morgan Stanley yields 5.14% ($t = 4.05$, SR = 1.22), Barclays 5.56% ($t = 3.62$, SR = 1.09), and RBC 3.15% ($t = 2.81$, SR = 0.71). Value-weighted consensus portfolios produce even larger returns at 8.14% ($t = 4.66$). Importantly, the strategy also performs well within the S&P 500 (4.88%, $t = 3.89$) and Russell 1000 (5.39%, $t = 5.82$) universes, where options are highly liquid and bid-ask spreads are narrow, suggesting that transaction costs are unlikely to erode the economic gains.

Panel B presents skewness-asset sorts. The consensus long-short portfolio earns 0.37% per month ($t = 3.24$, SR = 0.62). Morgan Stanley's signal is particularly strong, yielding 0.67% ($t = 2.61$, SR = 0.68).

Figures 2 and 3 plot 60-month rolling average returns and annualized Sharpe ratios for the straddle and skewness long-short portfolios within the S&P 500 universe, providing a visual assessment of how the analysts' higher moment expectations perform across different market regimes. Focusing on S&P 500 constituents ensures that the underlying options are highly liquid, so the documented profitability is not inflated by illiquidity premia or stale prices.

The straddle strategy accumulates gains steadily throughout the sample, with no prolonged drawdown period, indicating that the variance information embedded in analyst price targets is not concentrated in a single episode such as the financial crisis or the COVID-19 sell-off. The skewness strategy exhibits a similar pattern of broad-based profitability, albeit with smaller magnitudes consistent with the more subtle nature of higher-moment predictability. Both strategies generate positive returns in nearly all rolling windows, confirming that the economic value of analyst expectations is robust over time rather than an artifact of a particular subsample.

3.3. Decomposing Subjective Variance and Skewness

A distinctive feature of our setting is that the analyst's price target can often be decomposed into its underlying components. Many reports derive the target price as the product of a valuation multiple (e.g., a price-to-earnings ratio) and an earnings forecast (e.g., EPS) (Ben-David and Chinco, 2024), so that PT = Multiple × EPS. Taking logs, the analyst's expected return decomposes additively:

$$\log\left(\frac{\text{PT}}{P}\right) = \log(\text{Multiple}) + \log\left(\frac{\text{EPS}}{P}\right).$$

This additive structure in logs implies that the variance of the analyst's return expectation can be decomposed into a *valuation component*, $IV^{Multiple}$, capturing uncertainty in the multiple the analyst assigns across bull and bear scenarios, and an *earnings component*, IV^{EPS}, capturing uncertainty in the EPS forecast across scenarios. Intuitively, $IV^{Multiple}$ reflects the analyst's uncertainty about how the market will value the firm (e.g., through P/E expansion or compression), whereas IV^{EPS} reflects uncertainty about the firm's fundamental earnings trajectory. An

analogous decomposition applies to skewness.

To construct these components, we extract the numeric valuation multiples and EPS forecasts from each scenario: numeric multiples are available for approximately 74,000 reports and numeric EPS forecasts for approximately 30,000 reports. We then compute the variance and skewness of log(Multiple) and log(EPS) separately, using the same binomial probability $p = (\text{base} - \text{bear})/(\text{bull} - \text{bear})$ derived from the price targets.

Table 6 reports the results. Panel A examines realized variance and Panel B examines realized skewness. In each panel, column (1) replicates the baseline specification from Table 2 on the decomposition subsample, column (2) replaces the aggregate analyst expectation with the multiple component alone, and column (3) uses the EPS component alone.

For realized variance, Panel A column (2) shows that $IV^{Multiple}$ is a significant predictor (0.027, $t = 2.37$), indicating that analyst uncertainty about the valuation multiple carries information about future volatility. In contrast, IV^{EPS} in column (3) is positive but not statistically significant (0.016, $t = 1.45$), suggesting that either the sample size is too small to draw statistical inference or earnings uncertainty alone does not incrementally predict realized variance after controlling for option-implied and past realized variance. For realized skewness, both components are significant: $Skew^{Multiple}$ (0.052, $t = 3.25$) in Panel B column (2) and $Skew^{EPS}$ (0.047, $t = 2.14$) in column (3). These results suggest that the valuation channel is the primary driver of the predictive content for variance, while both valuation and earnings uncertainty contribute to predicting skewness. Having established that the predictive content of analyst expectations can be traced to specific economic components, we now examine whether sophisticated investors' derivative positions reflect these analyst signals before turning to narrative content.

3.4. Hedge Fund Derivative Ownership

If analyst variance expectations contain genuine information about future volatility, sophisticated investors may trade on this information in option markets. We test this hypothesis using institutional option ownership data from SEC 13F filings (see, Aragon, Chen, and Shi, 2022).

We focus on hedge funds (classified as alternative asset managers) and measure their call, put, and straddle ownership as a percentage of shares outstanding, where straddle ownership is defined as the minimum of call and put ownership to capture paired call-put positions.

Table 7 reports regressions of hedge fund derivative ownership on analyst variance expectations, controlling for option-implied variance, past realized variance, firm characteristics, and option-level controls, with firm and year-month fixed effects. The analyst variance expectation (IV^{Analyst}) is positively and significantly associated with hedge fund call ownership (column 1), put ownership (column 2), and straddle ownership (column 3). These results indicate that hedge funds increase both call and put positions, consistent with volatility trading, in stocks for which analyst expectations signal higher future variance. The significance of the straddle measure further suggests that these are not directional bets but rather paired positions that isolate volatility exposure, consistent with the straddle return predictability documented in Section 3.2.

3.5. Risk–Return and Skewness–Return Trade-Offs

Having established that analyst expectations are informative, we now ask whether they are consistent with classical asset pricing theories. Since each analyst report provides expectations for return, variance, and skewness simultaneously, we can test the risk–return and skewness–return trade-offs using a single agent's beliefs, avoiding the mismatch between subjective expectations and realized outcomes that plagues conventional tests. Specifically, we estimate the regression of the following form:

$$r^{\text{Analyst}}_{b,i,t\to t+12} = \alpha + \beta\, IV^{\text{Analyst}}_{b,i,t\to t+12} + \gamma\, \text{Controls}_{i,t} + \delta_b + \mu_t + \epsilon_{b,i,t},$$

$$r^{\text{Analyst}}_{b,i,t\to t+12} = \alpha + \beta\, Skew^{\text{Analyst}}_{b,i,t\to t+12} + \gamma\, \text{Controls}_{i,t} + \delta_b + \mu_t + \epsilon_{b,i,t},$$

where $r^{\text{Analyst}}_{b,i,t\to t+12} = \ln(\text{Price Target}/P_t)$ is the analyst's expected log return, b indexes brokers, i indexes firms, and t is the current stock price date. Controls follow Fama and French (2015): market beta (from a rolling one-year daily CAPM regression), log market capitalization

(log(ME)), log book-to-market (log(B/M)), asset growth (AG), and gross profitability-to-asset (GPA). δ_b and μ_t denote broker and year-month fixed effects. Standard errors are three-way clustered by broker, firm, and year-month. Independent variables are winsorized at the 1st and 99th percentiles.

Table 8 reports the results. Panel A studies the risk–return trade-off. In column (1), IV^{Analyst} is positively associated with the analyst's expected return (coefficient 0.282, $t = 4.79$). Columns (3)–(5) estimate the relationship separately for each of the three largest brokers: Barclays (0.172, $t = 2.03$), Morgan Stanley (0.078, $t = 3.12$), and RBC (0.388, $t = 20.27$). The positive trade-off is pervasive across firms.

Panel B examines the skewness–return trade-off. $Skew^{\text{Analyst}}$ enters negatively (coefficient -0.073, $t = -7.53$ in column 1), and the magnitude strengthens with controls (-0.088, $t = -11.01$ in column 2). The negative relationship holds across all three individual brokers: Barclays (-0.090, $t = -18.75$), Morgan Stanley (-0.134, $t = -19.82$), and RBC (-0.060, $t = -12.79$). These results are consistent with theories in which investors accept lower expected returns for positively skewed assets (Brunnermeier, Gollier, and Parker, 2007; Barberis and Huang, 2008; Mitton and Vorkink, 2007).

Taken together, these findings show that when both return and risk expectations come from the same agent, the data support a positive risk–return trade-off and a negative skewness–return trade-off, exactly as classical theories predict. This contrasts with the puzzling negative volatility–return relationship often documented using realized return data (Ang, Hodrick, Xing, and Zhang, 2006), and highlights the importance of using subjective expectations for both first-order and second-order moments simultaneously.

3.6. Determinants of Subjective Variance and Skewness

How does an analyst form expectations about higher moments? Understanding the expectation formation process is central to assessing whether analyst beliefs reflect systematic processing of publicly available information or incorporate private, qualitative judgments that go

beyond standard observable characteristics. If common firm characteristics and market-based moments fully explain analyst expectations, then the analyst's role is largely mechanical; if not, the residual variation points to proprietary information channels—such as the narrative content of analyst reports—that warrant further investigation.

To answer this question, we regress IV^{Analyst} and $Skew^{\text{Analyst}}$ on market-based moments and firm characteristics:

$$IV^{\text{Analyst}}_{b,i,t\to t+12} = \alpha + \beta_1\, IV^{\text{Option}}_{i,t\to t+12} + \beta_2\, RV^{\text{Past}}_{i,t-12\to t} + \gamma\, \text{Controls}_{i,t} + \delta_b + \mu_t + \epsilon_{b,i,t},$$

and analogously for $Skew^{\text{Analyst}}$ with option-implied skewness and past realized skewness as the key regressors. Controls follow the same specification as in the previous sections: market beta, log market capitalization (log(ME)), log book-to-market (log(B/M)), asset growth (AG), and gross profitability-to-asset (GPA). δ_b and μ_t denote broker and year-month fixed effects. Standard errors are three-way clustered by broker, firm, and year-month.

Panel A of Table 9 examines the determinants of IV^{Analyst}. In column (1), both IV^{Option} (0.434, $t = 10.15$) and past realized variance (0.303, $t = 4.33$) are strongly positively associated with IV^{Analyst}. The larger coefficient on IV^{Option} indicates that forward-looking market-implied variance plays a dominant role in shaping analysts' variance beliefs. Column (2) adds firm characteristics: log(B/M) enters negatively (-0.019, $t = -3.06$), and GPA is significantly negative (-0.170, $t = -3.43$), consistent with more profitable firms being perceived as less volatile. Market beta, log(ME), and AG are not significant after conditioning on IV^{Option} and past realized variance. Column (3) adds broker and year-month fixed effects with qualitatively similar results.

Panel B turns to $Skew^{\text{Analyst}}$. Option-implied skewness is strongly positive (0.342, $t = 10.63$ in column 1), while past realized skewness enters with a negative sign (-0.052, $t = -13.10$). The negative loading on past realized skewness implies a contrarian pattern: after experiencing high skewness, analysts expect a reversal. Among firm characteristics in column (2), log(ME) is the most prominent (-0.061, $t = -8.95$), indicating that analysts expect smaller firms to

have more positively skewed returns. GPA is positively associated with $Skew^{\text{Analyst}}$ (0.155, $t = 3.21$), and AG enters negatively (-0.054, $t = -3.20$). These patterns survive the inclusion of fixed effects in column (3).

A striking feature of Table 9 is how little explanatory power firm characteristics add beyond market-based moments. For subjective variance, including the full set of controls raises the adjusted R^2 from 0.326 to only 0.337; for subjective skewness, the improvement is from 0.043 to 0.067. Even after adding broker and year-month fixed effects, the adjusted R^2 for skewness reaches only 0.113. Much of the cross-sectional variation in analyst higher-moment expectations thus remains unexplained by standard observable characteristics and market-based moments alone. This suggests that analysts incorporate qualitative, firm-specific information that is not captured by traditional financial variables. To investigate these qualitative inputs, we next turn to the narrative content of analysts' scenario discussions.

4. Narratives of Analysts' Information Sets

The preceding sections establish that analyst expectations contain information beyond what is captured by market moments and standard firm characteristics. This raises a natural question: what information do analysts draw on when forming their higher-moment beliefs? To investigate, we exploit the textual discussions that accompany each analyst's bull and bear scenario forecasts.

4.1. Topic Extraction

A recent growing literature uses large language models (LLMs) to extract and structure the narrative content of analyst reports (see, e.g., Bastianello, Decaire, and Guenzel, 2024, 2025; Ke, 2024). Following this approach, we prompt an LLM to classify each scenario text into a topic taxonomy and use the resulting probability distributions to measure the allocation of analyst attention across different information categories. Intuitively, when an analyst devotes a larger share of her scenario discussion to a particular topic, the associated probability is higher,

reflecting greater attention to that factor in forming her expectation.[8]

Specifically, we develop a topic taxonomy through a multi-step procedure.[9] First, we use Gemini 2.0 Flash to perform open-ended topic extraction on each scenario text, allowing the model to identify all economically meaningful topics without imposing a pre-defined taxonomy. Second, we cluster the thousands of distinct raw topic labels into broader categories using text similarity and manual inspection. Third, we manually review and validate the resulting taxonomy, ensuring that each sub-topic is economically distinct and that the categories provide comprehensive coverage. This yields 32 sub-topics organized into three broad topic categories: 12 at the firm level (e.g., revenue/earnings growth, valuation multiples, profit margins, risk management), 10 at the industry level (e.g., regulation/policy, market share/competition, consumer demand), and 10 at the macro level (e.g., GDP/consumer spending, monetary policy, inflation).

We then use Gemini 2.0 Flash to classify each scenario text against the finalized taxonomy. The model assigns a two-level probability structure. At the first level, it assigns *broad topic-level probabilities*, the share of the scenario discussion devoted to firm-specific, industry, and macroeconomic factors, that sum to one. At the second level, it assigns *conditional sub-topic probabilities* within each broad topic category, also summing to one. Their product, the *joint probability*, measures each sub-topic's overall share of the scenario narrative and sums to one across all 32 sub-topics. The classification is performed separately for each scenario (bull, bear, and base), yielding a complete probability distribution over topics for each scenario in every report.

4.2. Analysts' Narrative Attention to Topics by Scenarios

Figure 4 displays average broad topic-level weights across scenarios. Firm-level topics command the largest share (76.8% overall), followed by industry (16.9%) and macroeconomy (6.3%). The allocation shifts meaningfully across scenarios: bear scenarios devote substantially

[8] Appendix A.2 provides full details on the topic extraction, taxonomy construction, manual validation, and classification prompt.

[9] see A.2 in the appendix for more details

more attention to industry (20.3% vs. 14.5% in bull) and macroeconomic topics (9.7% vs. 6.1% in bull), consistent with these risks becoming more salient when analysts contemplate downside outcomes.

Figure 5 presents average joint probabilities for each sub-topic. Revenue/earnings growth, valuation multiples, and risk management are the most prevalent firm-level sub-topics. Regulation/policy, market share/competition, and consumer demand lead the industry category. GDP/consumer spending, macro uncertainty, and currency/commodities are the most discussed macro topics.

Table 10 provides a detailed breakdown. Panel A reports broad topic-level weights: firm topics dominate across all three scenarios, but bear scenarios allocate a notably larger share to industry and macro factors, with a bear-minus-bull difference of +5.8% and +3.6%, respectively. Panel B reports conditional sub-topic weights within each broad topic category. Revenue/earnings growth is the dominant firm-level sub-topic (27.2%), followed by valuation multiples (25.6%) and risk management (15.7%). The bear–bull contrast is sharpest for risk management (+12.5%), which more than doubles its weight in bear relative to bull scenarios, highlighting its role in downside analysis. In the industry category, regulation/policy (20.6%) and market share/competition (18.9%) are the leading sub-topics. GDP/consumer spending (14.4%) and macro uncertainty (12.2%) lead the macro category, with macro uncertainty showing the largest bear–bull gap (+4.8%) among macro sub-topics.

The topics analysts discuss evolve over time, tracking changes in the macroeconomic environment. Figure 6 plots the joint probability of the top four macro-level sub-topics, GDP/consumer spending, macro uncertainty, currency/commodities, and geopolitical risk, under bear (red) and bull (green) scenarios. GDP/consumer spending attention in bear scenarios spikes during the Global Financial Crisis, peaks at its highest level around the European debt crisis, then gradually declines before rising again after COVID-19 and the Russia–Ukraine conflict; bear scenarios consistently receive higher weight than bull scenarios throughout the sample. Macro uncertainty follows a similar pattern, with bear-scenario attention peaking sharply around the European debt crisis and remaining elevated during the China slowdown, while bull-scenario

attention remains low, producing a persistently large bear–bull gap. Currency/commodities attention behaves differently: bear and bull scenarios track each other closely, with both rising during the trade war period and again around 2022, suggesting that commodity and currency factors feature symmetrically in upside and downside narratives. Geopolitical risk attention is more episodic, with bear-scenario spikes around the Global Financial Crisis and the European debt crisis, and a modest uptick following the Russia–Ukraine conflict. These time-series patterns confirm that analysts' narrative content responds to real-time macroeconomic developments rather than consisting of static boilerplate.

However, the topic weights measure *attention*, the share of narrative devoted to each topic, not the marginal impact of a topic on expectations. A topic that rarely appears in scenario discussions may nonetheless substantially shift an analyst's view when it does appear. For instance, an analyst may seldom discuss pandemic risk, but a single mention of COVID-19 could signal a large revision in her perceived tail risk. Linear regressions on the most-discussed topics cannot capture such effects, nor can they accommodate nonlinear interactions among the full set of 32 sub-topics. To identify which topics genuinely *drive* expectations, as opposed to merely correlating with them through attention, we turn to machine-learning methods which can accommodate nonlinear relationships and high-dimensional interactions (Gu, Kelly, and Xiu, 2020; Lundberg and Lee, 2017).

4.3. Machine-Learning Key Topics Driving Analysts' Expectations

We use Boosted Regression Trees (XGBoost) combined with SHAP values to quantify the importance of individual topics in explaining analyst expectations. For each report, we construct 64 topic features, the conditional sub-topic weights of all 32 sub-topics under both bull and bear scenarios, and combine them with market-based predictors (i.e., option-implied and past realized moments).[10]

[10]Boosted Regression Trees build an ensemble of shallow decision trees sequentially, with each tree trained to correct the residual errors of the preceding ensemble (Chen and Guestrin, 2016). This iterative structure naturally accommodates nonlinear relationships and high-dimensional interactions among features. Hyperparameters (tree depth and number of estimators) are selected via 5-fold cross-validated grid search on the training set. To obtain genuinely out-of-sample performance, we randomly hold out 20% of observations, stratified by broker, and

Table 11 reports out-of-sample R^2 from models that combine market moments with narrative topic features. The full topic model achieves an out-of-sample R^2 of 53.98% for variance and 29.25% for skewness, far exceeding the adjusted R^2 of 36.5% and 11.3% from Table 9, which rely on market moments, firm characteristics, and fixed effects. This large improvement confirms that analysts' narrative content captures information that standard observable characteristics miss, and that the modest R^2 from the determinants regression analysis reflects the absence of these qualitative inputs rather than noise in analyst expectations. Table 11 further decomposes the topic contribution by broad category. Among individual subsets, firm topics contribute the most incremental explanatory power (50.99% for variance, 23.93% for skewness), consistent with the broad topic-weight distribution in Figure 4. Industry and macro topics individually contribute less but still meaningfully exceed the market-moments-only baseline.

Figure 7 ranks features by mean absolute SHAP value for the variance model.[11] Option-implied variance and past realized variance are the two most important inputs overall. Among narrative variables, intangible capital, revenue/earnings, and valuation multiples rank highest, followed by profit margin, discount rate/WACC, and capital structure.

Figure 8 shows the corresponding ranking for skewness. While option-implied and realized skewness remain the dominant predictors, the mix of influential narrative topics shifts markedly: market share and profit margin rise to the top, followed by valuation multiples, credit lines, and COVID-19. Notably, the SHAP-based importance rankings differ from the attention rankings in Table 10: topics that receive modest narrative attention, such as intangible capital for variance or credit lines and COVID-19 for skewness, can nonetheless exert outsized influence on expectations, consistent with the distinction between attention and importance discussed above. This asymmetry in the topics driving variance versus skewness expectations reinforces the conclusion that higher moment expectations are shaped by distinct information sets and are not simply scaled versions of one another.

evaluate models on this held-out test set.

[11] SHAP values decompose each prediction into additive contributions from individual features, grounded in cooperative game theory (Lundberg and Lee, 2017). The mean absolute SHAP value of a feature measures its average impact on the model's output across all observations: a higher value indicates that the feature shifts predictions more, regardless of direction.

Figures A.4–A.6 in the Appendix display word clouds for each sub-topic, contrasting the vocabulary analysts use in bearish (red) versus bullish (green) scenarios. The clouds reveal that analysts draw on largely the same core terms across scenarios but shift emphasis in intuitive ways. For instance, within Revenue/Earnings Growth, bullish narratives foreground "upside" and "growth," whereas bearish narratives emphasize "downside" and "margin" compression. Several sub-topics also exhibit a pronounced volume asymmetry, with bearish scenarios generating denser word clouds than their bullish counterparts. This pattern is especially visible among macro-level topics such as Currency/Commodities, Inflation/Yield Curve, Monetary Policy/Rates, and Geopolitical Risk, where bearish narratives are dominated by terms like "oil," "rates," "interest," and "downside." A similar bear-heavy tilt appears in select firm- and industry-level topics, notably Credit Lines, Pricing/Input Costs, and Industry Utilization, suggesting that analysts devote more detailed narrative justification to downside risks than to upside opportunities when discussing cost pressures, financing conditions, and external shocks. These scenario-specific lexical patterns confirm that the topic classifications capture genuine differences in how analysts frame upside and downside risks, lending further support to the narrative-based features used in the machine-learning models.

Figure 9 synthesizes the foregoing results in a Sankey diagram that maps the flow of narrative influence from each sub-topic to variance and skewness expectations, with flow widths proportional to aggregated mean absolute SHAP values (combining bear and bull features). The diagram makes visually explicit that the two moments draw on overlapping yet distinct narrative inputs. Topics such as valuation multiples and profit margin send thick flows to both targets, whereas intangible capital and discount rate/WACC flow predominantly to variance and market share, credit lines, and COVID-19 flow disproportionately to skewness. Macro-level topics collectively account for a larger share of narrative influence on skewness than on variance, consistent with the view that tail-risk expectations are more sensitive to external shocks and aggregate conditions.

5. Additional Results and Robustness

5.1. Mixture Normal Model

Our baseline model infers subjective variance and skewness from a two-state binomial tree in which the stock price moves to the bull or bear price with probabilities p and $1-p$. This representation is exactly identified by the three scenario prices but imposes a discrete distribution with mass concentrated at two points. To assess the robustness of our findings to distributional assumptions, we consider a two-component mixture of normals that nests the binomial tree as a limiting case while allowing for continuous return variation within each state.

Specifically, the mixture density over log returns is:

$$f(r) = (1-w)\cdot\phi(r;\,\mu_{\text{bear}},\,\sigma^2_{\text{bear}}) + w\cdot\phi(r;\,\mu_{\text{bull}},\,\sigma^2_{\text{bull}}), \tag{2}$$

where $\phi(\cdot;\,\mu,\sigma^2)$ denotes the normal density with mean μ and variance σ^2. The component means are set to the log returns implied by the bear and bull prices, $\mu_{\text{bear}} = \ln(\text{Bear}/P_t)$ and $\mu_{\text{bull}} = \ln(\text{Bull}/P_t)$, and the mixture weight w equals the binomial probability $p = (\text{Target} - \text{Bear})/(\text{Bull} - \text{Bear})$. Unlike a symmetric specification with a single shared σ, we allow the within-component standard deviations to differ across the bear and bull states:

$$\sigma_{\text{bear}} = \lambda\cdot(r_{\text{base}} - \mu_{\text{bear}}), \qquad \sigma_{\text{bull}} = \lambda\cdot(\mu_{\text{bull}} - r_{\text{base}}), \tag{3}$$

where $r_{\text{base}} = \ln(\text{Target}/P_t)$ is the base-case log return. Under this parameterization, each component's standard deviation is proportional to its distance from the base case, so that scenarios farther from the analyst's central estimate carry more within-state uncertainty. The single free parameter $\lambda \geq 0$ is calibrated so that the median of the mixture distribution coincides with the base-case log return.[12]

[12]The mixture CDF at r_{base} simplifies to $F(r_{\text{base}}) = (1-w)\,\Phi(1/\lambda) + w\,\Phi(-1/\lambda)$, which depends on λ only through the standardized argument $1/\lambda$. At $\lambda = 0$, $F = 1-w$; as $\lambda \to \infty$, $F \to 0.5$. Bisection converges in 60 iterations across all 180,858 reports in under one second. We cap λ at 0.5 to ensure the within-component standard deviations do not exceed half the distance between each scenario and the base case.

The asymmetric specification is important because it generates a skewness term absent from the binomial tree. In the data, the bear scenario is typically farther from the base case than the bull scenario: 89% of reports have $r_{\text{base}} - \mu_{\text{bear}} > \mu_{\text{bull}} - r_{\text{base}}$, with a mean ratio of 2.5. Since σ_{bear} and σ_{bull} are proportional to these distances, the bear component receives more dispersion than the bull component, producing a heavier left tail.

The mixture model has closed-form moments. The variance decomposes into within-component and between-component terms:

$$IV^{\text{Mixture}} = \underbrace{(1-w)\,\sigma^2_{\text{bear}} + w\,\sigma^2_{\text{bull}}}_{\text{within-component}} + \underbrace{w(1-w)\,\delta^2}_{=IV^{\text{Analyst}}}, \tag{4}$$

where $\delta = \mu_{\text{bull}} - \mu_{\text{bear}}$. The third central moment is:

$$M_3^{\text{Mixture}} = \underbrace{w(1-w)(1-2w)\,\delta^3}_{\text{binomial component}} + \underbrace{3\,w(1-w)\,(\sigma^2_{\text{bull}} - \sigma^2_{\text{bear}})\,\delta}_{\text{asymmetric component}}. \tag{5}$$

The first term is the binomial tree's third moment. The second term is the new contribution from asymmetric variances: when $\sigma_{\text{bear}} > \sigma_{\text{bull}}$, it is negative, adding left-tail skewness beyond what the binomial captures. In the data, this asymmetric component dominates: the mean asymmetric contribution to M_3 is -0.085, while the binomial component contributes $+0.013$. The asymmetric term thus reverses the sign and accounts for more than 100% of the total third moment ($M_3 = -0.071$). Standardized skewness is:

$$Skew^{\text{Mixture}} = \frac{M_3^{\text{Mixture}}}{\left(IV^{\text{Mixture}}\right)^{3/2}}. \tag{6}$$

When $\lambda = 0$, the within-component variances vanish, and the mixture collapses to the binomial tree.

The resulting σ_{bear} (mean 0.240) exceeds σ_{bull} (mean 0.119) by a factor of two, reflecting the typical asymmetry in scenario placement. The mixture variance (IV^{Mixture}, mean 0.239) exceeds the binomial variance (mean 0.187) by the within-component contribution, while the

mixture skewness (mean −0.616) is more negative than the binomial (mean −0.357), amplified by the asymmetric variance term. Cross-sectional rankings remain highly preserved: Spearman rank correlations between binomial and mixture measures are 0.998 for variance and 0.989 for skewness. Notably, the within-model correlation between variance and skewness drops from 0.227 under the binomial to 0.121 under the mixture, suggesting that the asymmetric σ component generates skewness variation that is not mechanically tied to variance the way the binomial's third moment is.

Table A.2 replicates the realized moment forecasting regressions of Table 2 using the mixture moments. In Panel A, IV^{Mixture} alone predicts future realized variance with a coefficient of 0.428 ($t = 12.93$). In the most demanding specification with broker×year-month fixed effects (column 4), the coefficient is 0.122 ($t = 8.50$), closely matching the binomial model's 0.159 ($t = 8.33$) in Table 2. The coefficients on the option-implied and past realized controls are virtually identical across the two models, and the adjusted R^2 is the same at 0.382. Panel B turns to skewness. In column (1), $Skew^{\text{Mixture}}$ predicts future realized skewness with a coefficient of 0.073 ($t = 8.29$). Notably, in the saturated specification (column 4), $Skew^{\text{Mixture}}$ remains the sole significant predictor (coefficient 0.063, $t = 4.32$), while option-implied and past realized skewness remain statistically insignificant. This result parallels the baseline finding that analyst scenario-based expectations are the only robust predictor of future realized skewness, and the mixture model's significance is comparable to the binomial's (0.075, $t = 4.12$).

Table A.3 replicates the risk–return and skewness–return trade-off regressions of Table 8. Panel A shows that higher IV^{Mixture} is associated with higher expected returns, with a positive and significant coefficient across all specifications, including when estimated separately for the three largest brokers. Panel B confirms the negative skewness–return relationship: stocks with more negative $Skew^{\text{Mixture}}$ command higher expected returns. These results are qualitatively consistent with the baseline, indicating that the subjective risk–return and skewness–return trade-offs are robust to the distributional assumption underlying moment extraction.

Taken together, the mixture model results demonstrate that the informational content of analyst scenario-based forecasts is not an artifact of the discrete two-state binomial parameteriza-

tion. Allowing for continuous, asymmetric return variation within each scenario state preserves the key empirical findings: predictability of realized moments and the classical risk–return and skewness–return trade-offs.

5.2. Three-State Model

Our binomial tree model assigns probabilities p and $1-p$ to the bull and bear scenarios, with p exactly identified from the three scenario prices. A limitation of this two-state framework is that bull and bear probabilities are mechanically linked—any increase in p decreases $1-p$—which generates an artificial positive correlation between implied variance and skewness. We address this concern by developing a three-state model in which the bull, bear, and base scenarios each receive separate probability mass.

Specifically, we assign probability p_{up} to the bull return $r_{up} = \ln(\text{Bull}/P_t)$, probability p_{down} to the bear return $r_{down} = \ln(\text{Bear}/P_t)$, and probability $p_{base} = 1 - p_{up} - p_{down}$ to the base return $r_{base} = \ln(\text{Target}/P_t)$. Unlike the binomial tree, the constraint $p_{up} + p_{down} = 1$ is relaxed, allowing the base case to carry positive probability mass. The model has two unknowns (p_{up} and p_{down}) and is identified by two moment conditions.

The first condition requires that the probability-weighted expected price equals the analyst's base-case price:

$$p_{up} \cdot \mathbb{E}[a] = p_{down} \cdot \mathbb{E}[b], \tag{7}$$

where $a_i = r_{up,i} - r_{base,i}$ is the upside deviation and $b_i = r_{base,i} - r_{down,i}$ is the downside deviation for report i. This pins down the ratio $p_{up}/p_{down} = \mathbb{E}[b]/\mathbb{E}[a]$, but a second condition is needed to determine the levels.

The second condition imposes that analysts correctly forecast variance on average:

$$\mathbb{E}_i\left[p_{up} \cdot a_i^2 + p_{down} \cdot b_i^2\right] = \mathbb{E}_i\left[RV_i^{Realized}\right]. \tag{8}$$

This assumption reflects the view that, although individual forecasts may be noisy, a broker's

analysts collectively have the right mental model for variance—the probability weights that rationalize their scenario forecasts are consistent with realized return variation on average. Combining the two conditions yields a closed-form solution:

$$p_{down} = \frac{S}{\frac{B}{A} \cdot C + D}, \qquad p_{up} = p_{down} \cdot \frac{B}{A}, \tag{9}$$

where $A = \mathbb{E}[a]$, $B = \mathbb{E}[b]$, $C = \mathbb{E}[a^2]$, $D = \mathbb{E}[b^2]$, and $S = \mathbb{E}[RV^{Realized}]$. Expectations are taken over all reports within a broker.

We estimate (p_{up}, p_{down}) separately for each broker using an expanding window of past reports. To avoid look-ahead bias, we use reports from years $\leq t - 2$ to estimate probabilities applied to year-t reports. This two-year gap ensures no overlap between the forward-looking returns used in estimation (which extend at most into year $t-1$) and the forward-looking returns of reports in year t. We require at least 50 historical reports per broker-year for estimation.

Given the estimated probabilities, the three-state moments are:

$$IV^{3State} = p_{up} \cdot a^2 + p_{down} \cdot b^2, \tag{10}$$

$$Skew^{3State} = \frac{p_{up} \cdot a^3 - p_{down} \cdot b^3}{\left(IV^{3State}\right)^{3/2}}. \tag{11}$$

The estimated probabilities are $p_{up} \approx 0.61$, $p_{down} \approx 0.32$, and $p_{base} \approx 0.06$, indicating that the base case receives a modest but nonzero probability weight. Importantly, the three-state model substantially reduces the mechanical coupling between variance and skewness: the within-model correlation between IV^{3State} and $Skew^{3State}$ is -0.05, compared with $+0.23$ for the binomial model. This decorrelation arises because allowing independent variation in p_{up} and p_{down} eliminates the constraint that all probability redistribution between the two tails must be zero-sum.

Despite this structural difference, the three-state variance is highly correlated with the binomial variance (correlation 0.92), confirming that the cross-sectional ranking of firms by subjective uncertainty is robust to the probability structure. The three-state variance also ex-

hibits modestly stronger forecasting power for option-implied variance (correlation 0.558 versus 0.544) and realized variance (correlation 0.412 versus 0.401).

Table A.4 replicates the realized moment forecasting regressions of Table 2 using the three-state moments. In Panel A, IV^{3State} predicts future realized variance with magnitudes and significance levels comparable to the baseline binomial specification. The option-implied and past realized variance controls carry similar coefficients, and the adjusted R^2 is virtually identical. Panel B examines skewness forecasting. $Skew^{3State}$ remains a significant predictor of future realized skewness, and the three-state skewness retains its predictive content even in the most saturated specification with broker×year-month fixed effects, paralleling the baseline finding that analyst-derived skewness expectations are the only robust predictor of future realized skewness.

Table A.5 replicates the risk–return and skewness–return trade-off regressions of Table 8. Panel A confirms the positive risk–return trade-off: higher IV^{3State} is associated with higher analyst return expectations across all specifications, including broker-level regressions. Panel B examines the skewness–return relationship. Stocks with more negative $Skew^{3State}$ command higher expected returns, consistent with the theoretical prediction that investors require compensation for bearing left-tail risk.

These results demonstrate that the key empirical findings—predictability of realized moments, the risk–return trade-off, and the skewness–return trade-off—are robust to relaxing the two-state probability structure. The three-state model offers the additional advantage of decoupling variance and skewness variation, providing a cleaner identification of each moment's distinct asset pricing content.

6. Conclusion

This paper constructs the first firm-level measures of subjective variance and skewness expectations by calibrating a binomial tree model to bull, base, and bear price targets from analyst equity research reports across the largest ten brokerage firms over 2007–2024. We document

three main findings. First, analyst variance and skewness expectations are informative; they positively predict future realized moments and cross-sectional option returns incrementally over option-implied and past realized moments. Second, exploiting the joint expectations for return, variance, and skewness from the same analyst, we recover a positive risk–return and a negative skewness–return trade-off, consistent with the predictions of Merton (1987), Brunnermeier, Gollier, and Parker (2007), and Barberis and Huang (2008). Third, using large language models to classify scenario narratives and machine learning to quantify topic importance, we show that analysts' qualitative assessments of firm, industry, and macroeconomic conditions are the key driver of their higher moment expectations.

These findings carry important implications. The informativeness of analyst higher-moment expectations suggests that scenario-based analysis disciplines risk assessment by requiring analysts to consider multiple states of the world. Level biases present in bull, base, and bear targets may partially cancel when computing central moments such as variance and skewness. The subjective risk–return trade-offs we document provide direct support for classical theories, underscoring the importance of distinguishing between subjective beliefs and realized outcomes in asset pricing tests. Our framework for linking narrative content to subjective expectations through LLMs and SHAP-based importance analysis offers a general methodology applicable to any setting where narratives shape investors' beliefs.

Several directions for future research emerge. As scenario-based frameworks become more widely adopted, extending our methodology to international markets would allow researchers to study important questions such as geoeconomic risks, a prominent macro-level topic in scenario narratives. Investigating the real-time dynamics of how analysts update their higher-moment beliefs in response to news events could also be fruitful. Understanding how such updates propagate to option prices would shed further light on the channels through which subjective expectations are incorporated into asset prices.

REFERENCES

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang, 2006, "The cross-section of volatility and expected returns," *The journal of finance*, 61(1), 259–299.

Aragon, G. O., and S. Chen, 2024, "Machine-learning about ESG preferences: Evidence from fund flows," *Available at SSRN 4522156*.

Aragon, G. O., S. Chen, and Z. Shi, 2022, "Volatility timing using etf options: Evidence from hedge funds," *Available at SSRN 4246146*.

Bakshi, G., N. Kapadia, and D. Madan, 2003, "Stock return characteristics, skew laws, and the differential pricing of individual equity options," *The Review of Financial Studies*, 16(1), 101–143.

Bali, T. G., and S. Murray, 2013, "Does risk-neutral skewness predict the cross-section of equity option portfolio returns?," *Journal of Financial and Quantitative Analysis*, 48(4), 1145–1171.

Barberis, N., and M. Huang, 2008, "Stocks as lotteries: The implications of probability weighting for security prices," *American Economic Review*, 98(5), 2066–2100.

Bastianello, F., P. Decaire, and M. Guenzel, 2024, "Mental models and financial forecasts," *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

——— , 2025, "Valuation models," *Working Paper*.

Ben-David, I., and A. Chinco, 2024, "Expected eps× trailing p/e," working paper, National Bureau of Economic Research.

Ben-David, I., J. R. Graham, and C. R. Harvey, 2013, "Managerial miscalibration," *The Quarterly journal of economics*, 128(4), 1547–1584.

Bollerslev, T., G. Tauchen, and H. Zhou, 2009, "Expected stock returns and variance risk premia," *The Review of Financial Studies*, 22(11), 4463–4492.

Brav, A., and R. Lehavy, 2003, "An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics," *The Journal of Finance*, 58(5), 1933–1967.

Brunnermeier, M. K., C. Gollier, and J. A. Parker, 2007, "Optimal beliefs, asset prices, and the preference for skewed returns," *American Economic Review*, 97(2), 159–165.

Campbell, J. Y., and T. Ramadorai, 2026, "Household Finance in Retrospect and Prospect," Working Paper 34621, National Bureau of Economic Research.

Cao, J., and B. Han, 2013, "Cross section of option returns and idiosyncratic stock volatility," *Journal of Financial Economics*, 108(1), 231–249.

Cao, J., B. Han, G. Li, R. Yang, and X. Zhan, 2024, "Forecasting option returns with news," *Available at SSRN*.

Cao, S., W. Jiang, J. Wang, and B. Yang, 2024, "From Man vs. Machine to Man+ Machine: The Art and AI of Stock Analyses," *Journal of Financial Economics*, 160, 103910.

Carr, P., and L. Wu, 2009, "Variance risk premiums," *The Review of Financial Studies*, 22(3), 1311–1341.

Chen, S., and S. Li, 2025, "Hedge Fund Option Demand and Crash Risk Premium," .

Chen, S., L. Peng, and D. Zhou, 2024, "Wisdom or Whims? Decoding Investor Trading Strategies with Large Language Models," *Available at SSRN 4867401*.

Chen, T., and C. Guestrin, 2016, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

De Silva, T., and D. Thesmar, 2024, "Noise in Expectations: Evidence from Analyst Forecasts," *The Review of Financial Studies*, 37(5), 1494–1537.

Fama, E. F., and K. R. French, 2015, "A five-factor asset pricing model," *Journal of financial economics*, 116(1), 1–22.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, return, and equilibrium: Empirical tests," *Journal of political economy*, 81(3), 607–636.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus, 2021, "Five facts about beliefs and portfolios," *American Economic Review*, 111(5), 1481–1522.

Goetzmann, W. N., D. Kim, and R. J. Shiller, 2022, "Crash narratives," working paper, National Bureau of Economic Research.

Goyal, A., and A. Saretto, 2009, "Cross-section of option returns and volatility," *Journal of Financial Economics*, 94(2), 310–326.

Gu, S., B. Kelly, and D. Xiu, 2020, "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies*, 33(5), 2223–2273.

Harvey, C. R., and A. Siddique, 2000, "Conditional skewness in asset pricing tests," *The Journal of finance*, 55(3), 1263–1295.

Heston, S. L., C. S. Jones, M. Khorram, S. Li, and H. Mo, 2023, "Option momentum," *The Journal of Finance*, 78(6), 3141–3192.

Hirshleifer, D., 2020, "Presidential Address: Social Transmission Bias in Economics and Finance," *Journal of Finance*, 75(4), 1779–1831.

Horenstein, A. R., A. Vasquez, and X. Xiao, 2023, "Common factors in equity option returns," *Available at SSRN 3290363*.

Jensen, T. I., 2023, "Subjective risk and return," *Available at SSRN 4276760*.

Jha, M., J. Qian, M. Weber, and B. Yang, 2024, "ChatGPT and Corporate Policies," working paper, National Bureau of Economic Research.

Jiang, J., B. T. Kelly, and D. Xiu, 2024, "Expected Returns and Large Language Models," *Available at SSRN*.

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

Ke, S., 2024, "Analysts' belief formation in their own words," *Available at SSRN 5025830*.

Kozhan, R., A. Neuberger, and P. Schneider, 2013, "The skew risk premium in the equity index market," *The Review of Financial Studies*, 26(9), 2174–2203.

Langlois, H., 2020, "Measuring skewness premia," *Journal of Financial Economics*, 135(2), 399–424.

Li, K., F. Mai, R. Shen, C. Yang, and T. Zhang, 2023, "Dissecting Corporate Culture Using Generative AI–Insights from Analyst Reports," *Available at SSRN 4558295*.

Lochstoer, L. A., and T. Muir, 2022, "Volatility expectations and returns," *The Journal of Finance*, 77(2), 1055–1096.

Lopez-Lira, A., and Y. Tang, 2023, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," *arXiv preprint arXiv:2304.07619*.

Lundberg, S. M., and S.-I. Lee, 2017, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, 30.

Manela, A., and A. Moreira, 2017, "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123(1), 137–162.

Merton, R. C., 1987, "A simple model of capital market equilibrium with incomplete information," .

Mitton, T., and K. Vorkink, 2007, "Equilibrium underdiversification and the preference for skewness," *The Review of Financial Studies*, 20(4), 1255–1288.

Nagel, S., and Z. Xu, 2022, "Asset pricing with fading memory," *The Review of Financial Studies*, 35(5), 2190–2245.

——— , 2023, "Dynamics of subjective risk premia," *Journal of Financial Economics*, 150(2), 103713.

Orlowski, P., P. Schneider, and F. Trojani, 2023, "On the nature of (jump) skewness risk premia," *Management Science*.

Shiller, R. J., 2017, "Narrative Economics," *American Economic Review*, 107(4), 967–1004.

Vasquez, A., 2017, "Equity volatility term structures and the cross section of option returns," *Journal of Financial and Quantitative Analysis*, 52(6), 2727–2754.

Weyns, G., J.-L. Perez, B. Hurewitz, and V. Jenkins, 2011, "Morgan Stanley's Risk-Reward Views: Unlocking the Full Potential of Fundamental Analysis," *Journal of Applied Corporate Finance*, 23(2), 59–68.

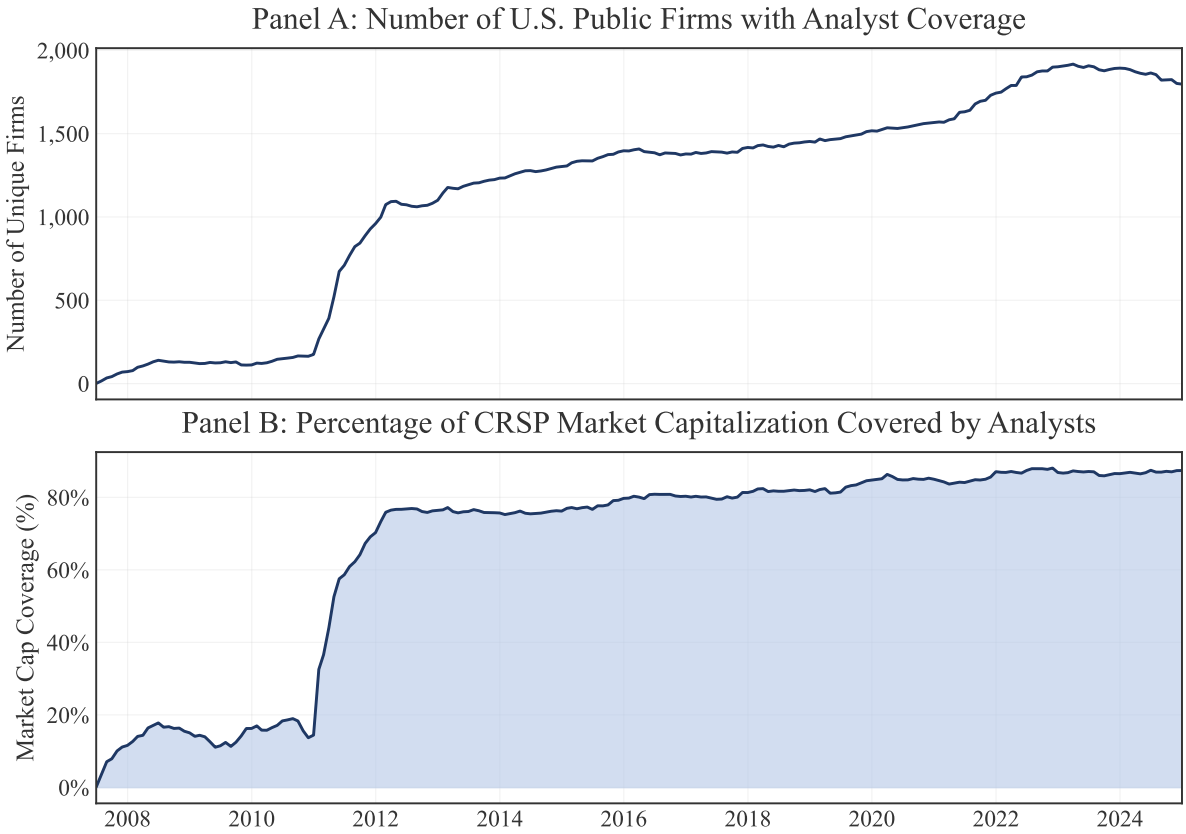


Fig. 1. **U.S. Public Firm Coverage by Analyst Reports**

This figure plots the monthly breadth of our analyst scenario-based sample over June 2007 to December 2024. A firm is classified as covered in month t if at least one brokerage firm issued a scenario-based report for that firm within the preceding 12 months. Panel A reports the number of unique firms with active coverage. Panel B reports the percentage of total CRSP common-stock market capitalization accounted for by covered firms.

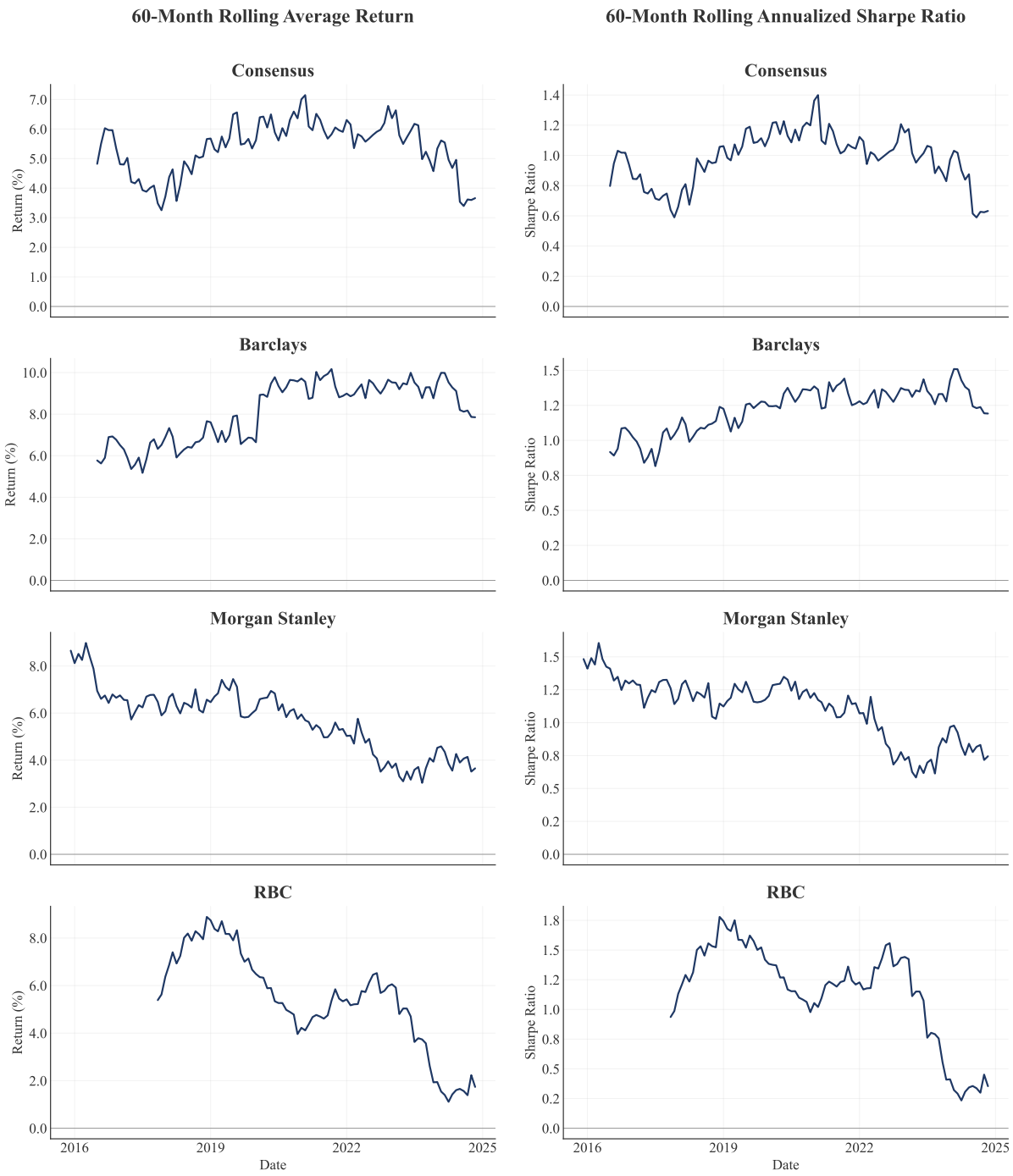


Fig. 2. **Performance of Long-Short Straddle Strategy on S&P 500 Firms**

This figure plots the 60-month rolling average return (left column) and rolling annualized Sharpe ratio (right column) of the equally-weighted long-short straddle portfolio sorted on orthogonalized analyst subjective variance (IV^{Analyst}) within the S&P 500 universe. Each month, the analyst signal is orthogonalized with respect to the full set of controls in Table 4. The portfolio buys the top quintile and sells the bottom quintile. Rows correspond to the consensus signal and the three largest individual brokers: Barclays, Morgan Stanley, and RBC.

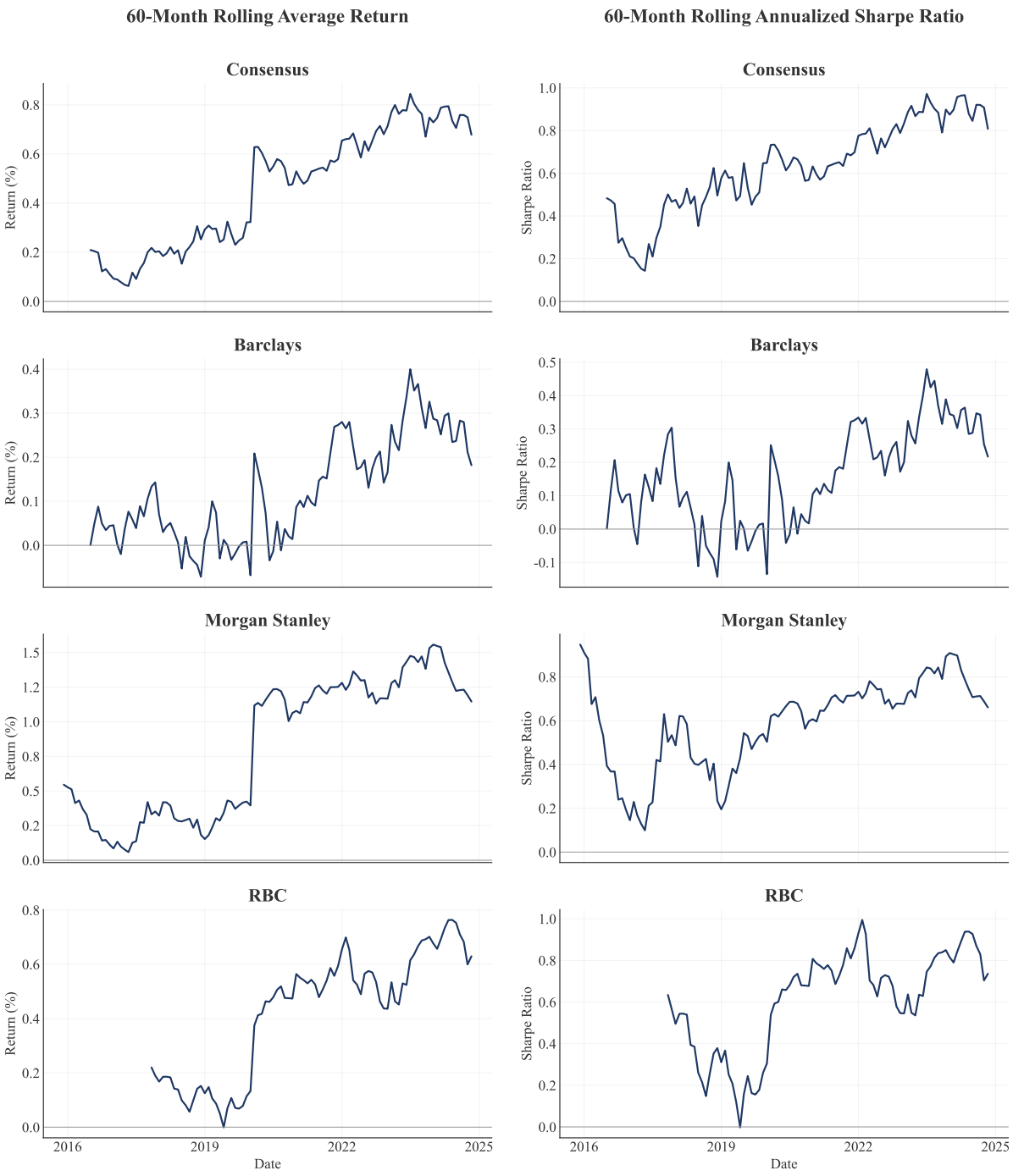


Fig. 3. **Performance of Long-Short Skewness Strategy on S&P 500 Firms**

This figure plots the 60-month rolling average return (left column) and rolling annualized Sharpe ratio (right column) of the equally-weighted long-short skewness-asset portfolio sorted on orthogonalized analyst subjective skewness ($Skew^{\text{Analyst}}$) within the S&P 500 universe. Each month, the analyst signal is orthogonalized with respect to the full set of controls in Table 4. The portfolio buys the top quintile and sells the bottom quintile. Rows correspond to the consensus signal and the three largest individual brokers: Barclays, Morgan Stanley, and RBC.

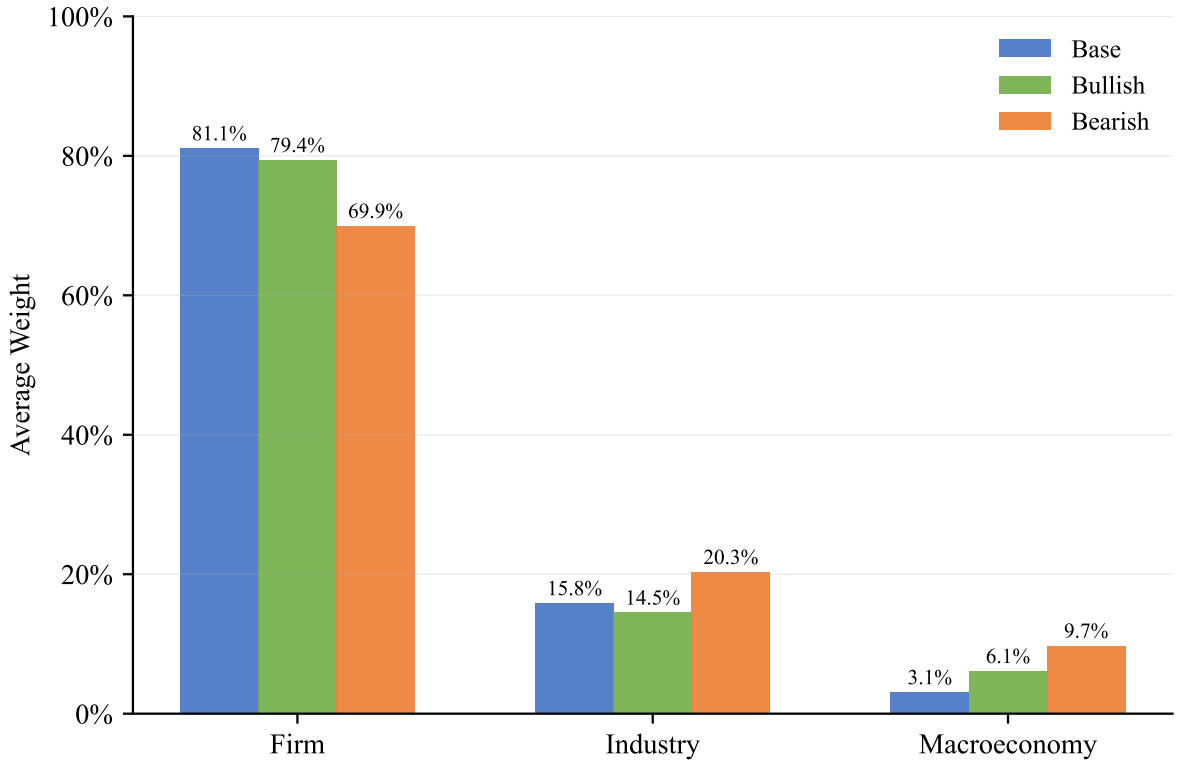


Fig. 4. **Broad Topic-Level Weights by Scenario**

This figure displays the average broad topic-level weights—firm, industry, and macroeconomy—across base, bull, and bear scenarios. Broad topic weights represent the fraction of the scenario text attributed to each category and sum to one within each scenario. The sample focuses on Barclays, Morgan Stanley, and RBC in which scenario narratives are available in analyst reports.

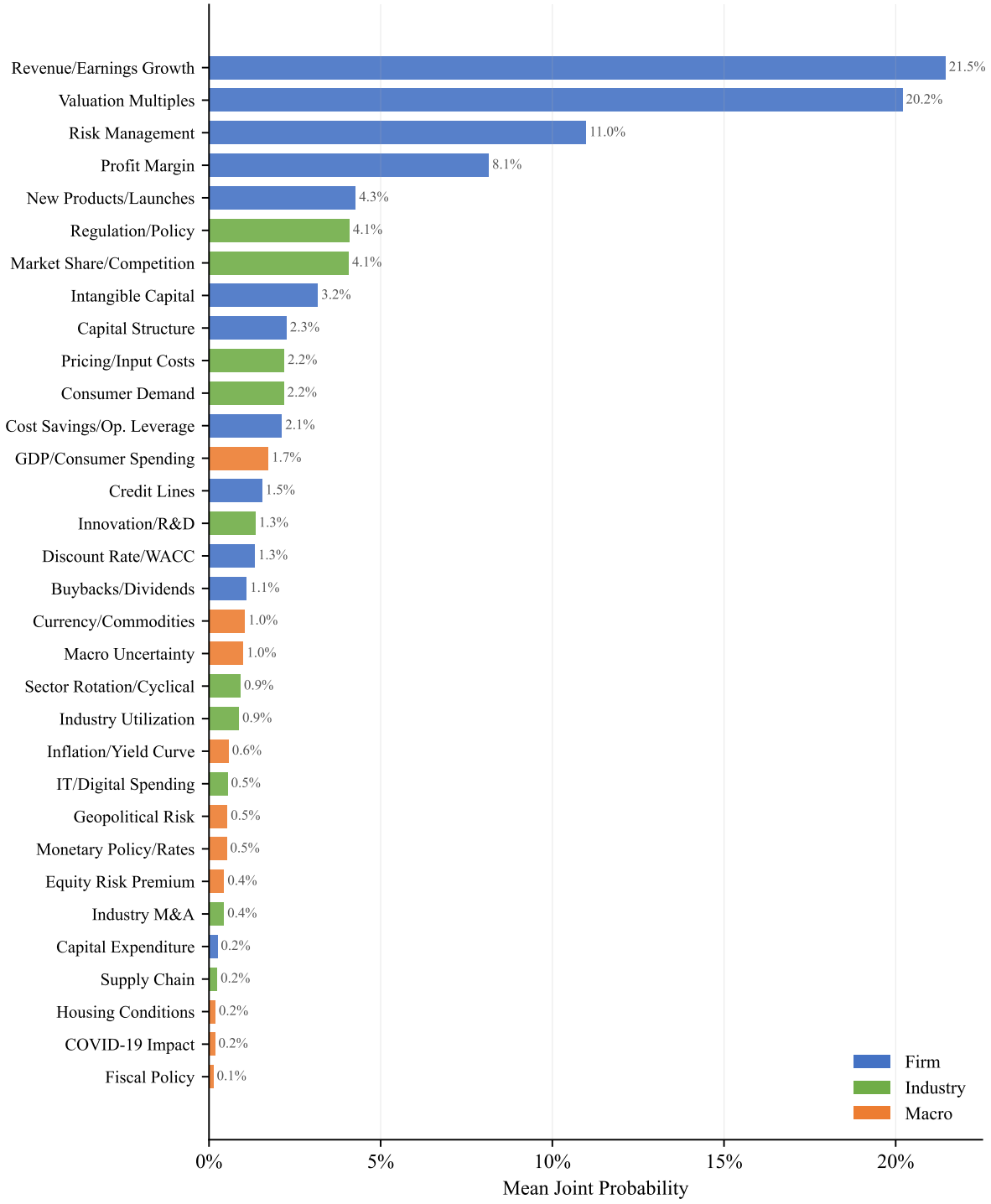


Fig. 5. **Sub-Topic Prevalence Across Scenarios**

This figure displays the average joint probability of each sub-topic across base, bull, and bear scenarios. Joint probability is the product of the broad topic-level weight and the conditional sub-topic weight, representing each sub-topic's share of the total scenario text. Sub-topics are grouped by broad topic category (firm, industry, macroeconomy) and sorted by overall prevalence. The sample focuses on Barclays, Morgan Stanley, and RBC in which scenario narratives are available in analyst reports.

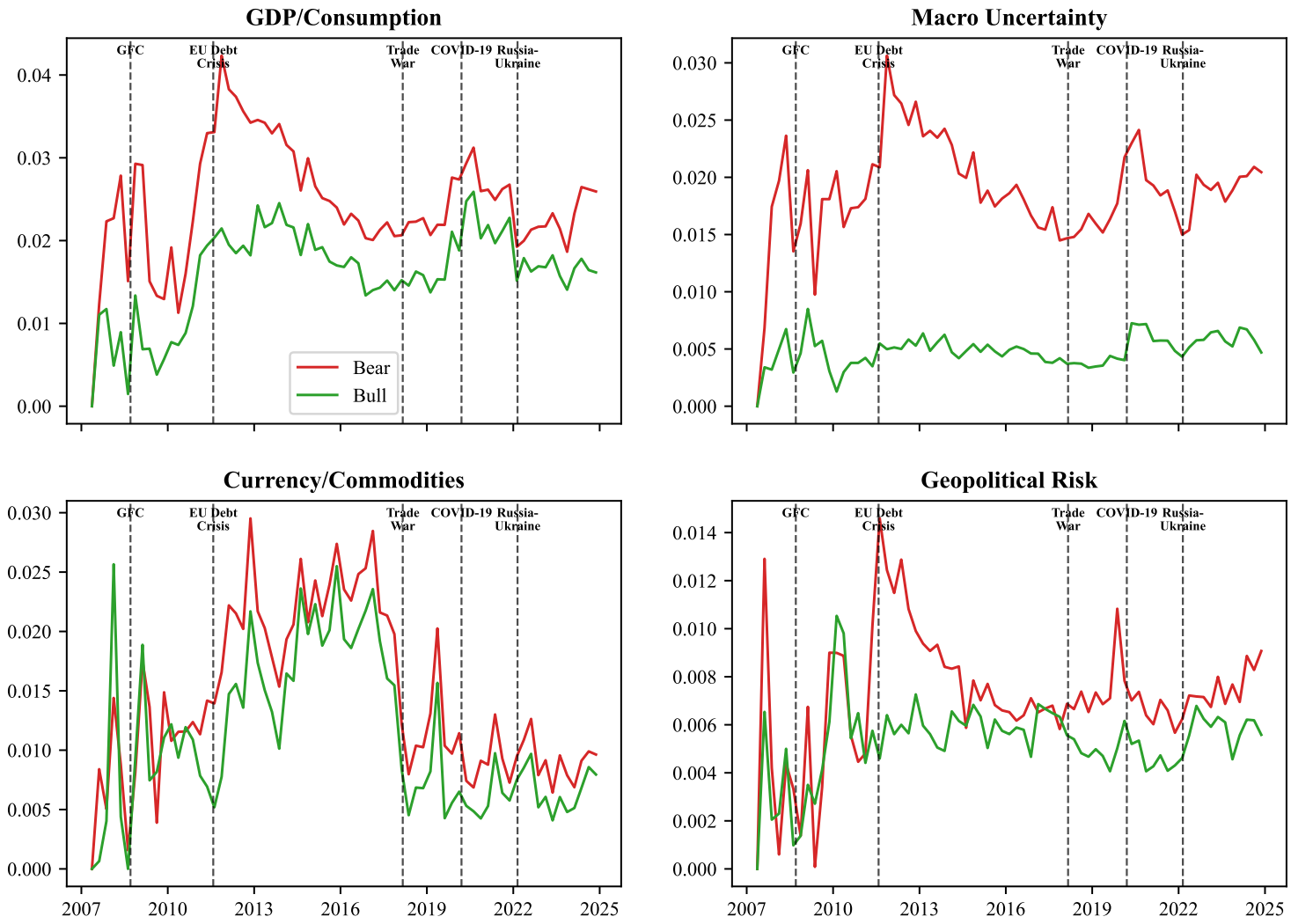


Fig. 6. **Time Variation in Macro Topic Focus**

This figure plots the quarterly average joint probability of the top four macro-level sub-topics across bear (red) and bull (green) scenarios. Joint probability is the product of the macro-level weight and the conditional sub-topic weight, representing each sub-topic's share of the total scenario text. Dashed vertical lines mark key macroeconomic events. The sample focuses on Barclays, Morgan Stanley, and RBC in which scenario narratives are available in analyst reports.

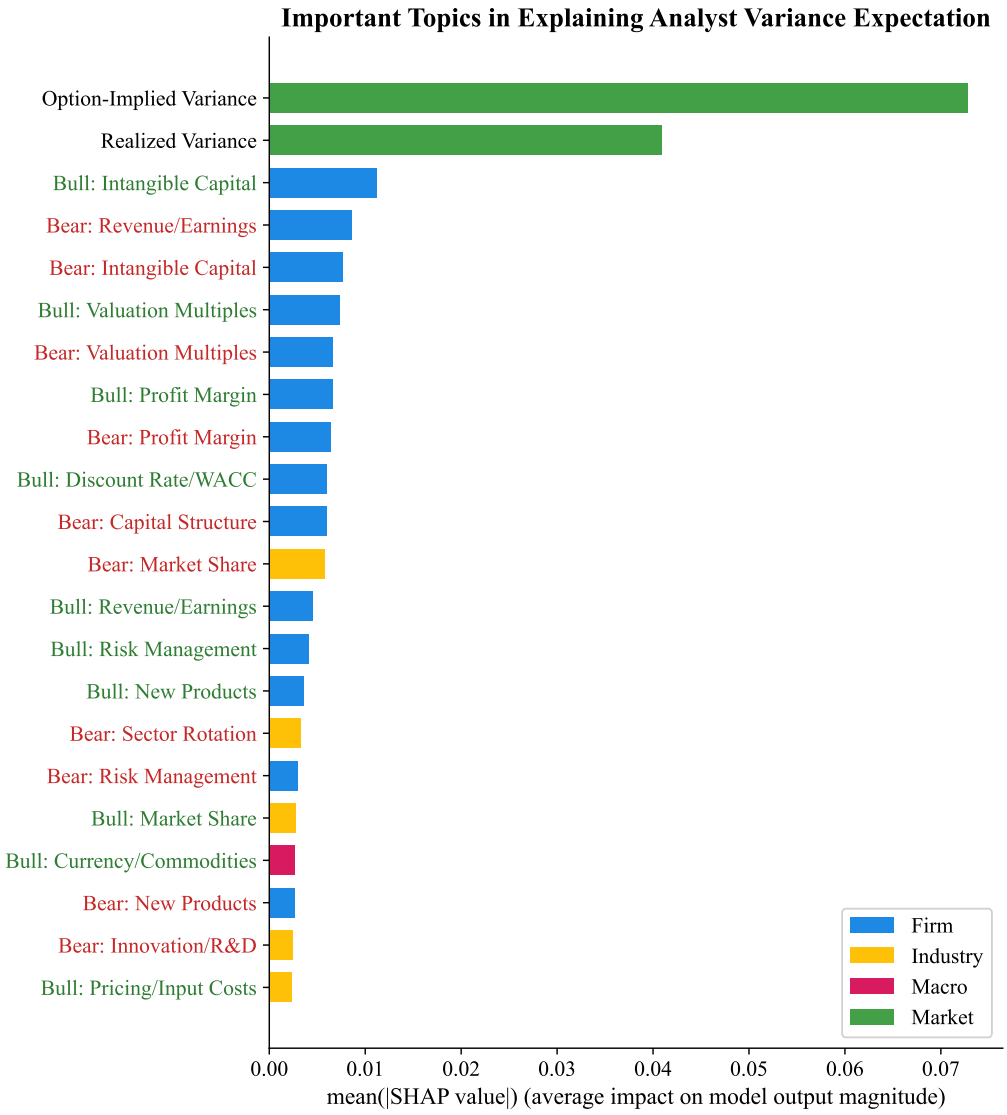


Fig. 7. **SHAP Feature Importance: Analyst Variance Expectation.**

This figure presents the mean absolute SHAP values from an XGBoost model predicting analyst variance expectations. Bars are colored by entity category: firm (blue), industry (amber), macro (pink), and market-based predictors (green). The sample includes analyst reports from Barclays, Morgan Stanley, and RBC over 2007–2024.

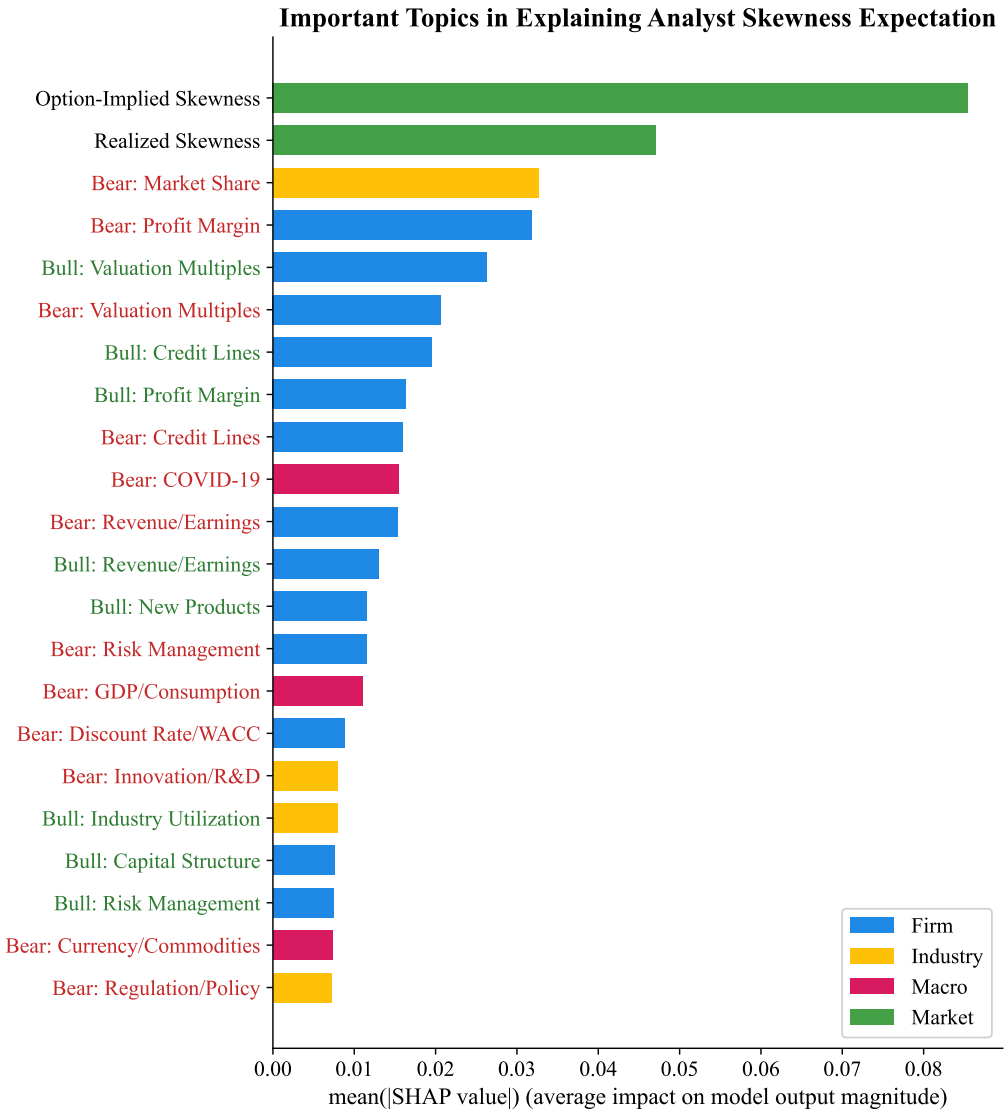


Fig. 8. **SHAP Feature Importance: Analyst Skewness Expectation.**

This figure presents the mean absolute SHAP values from an XGBoost model predicting analyst skewness expectations. Bars are colored by entity category: firm (blue), industry (amber), macro (pink), and market-based predictors (green). The sample includes analyst reports from Barclays, Morgan Stanley, and RBC over 2007–2024.

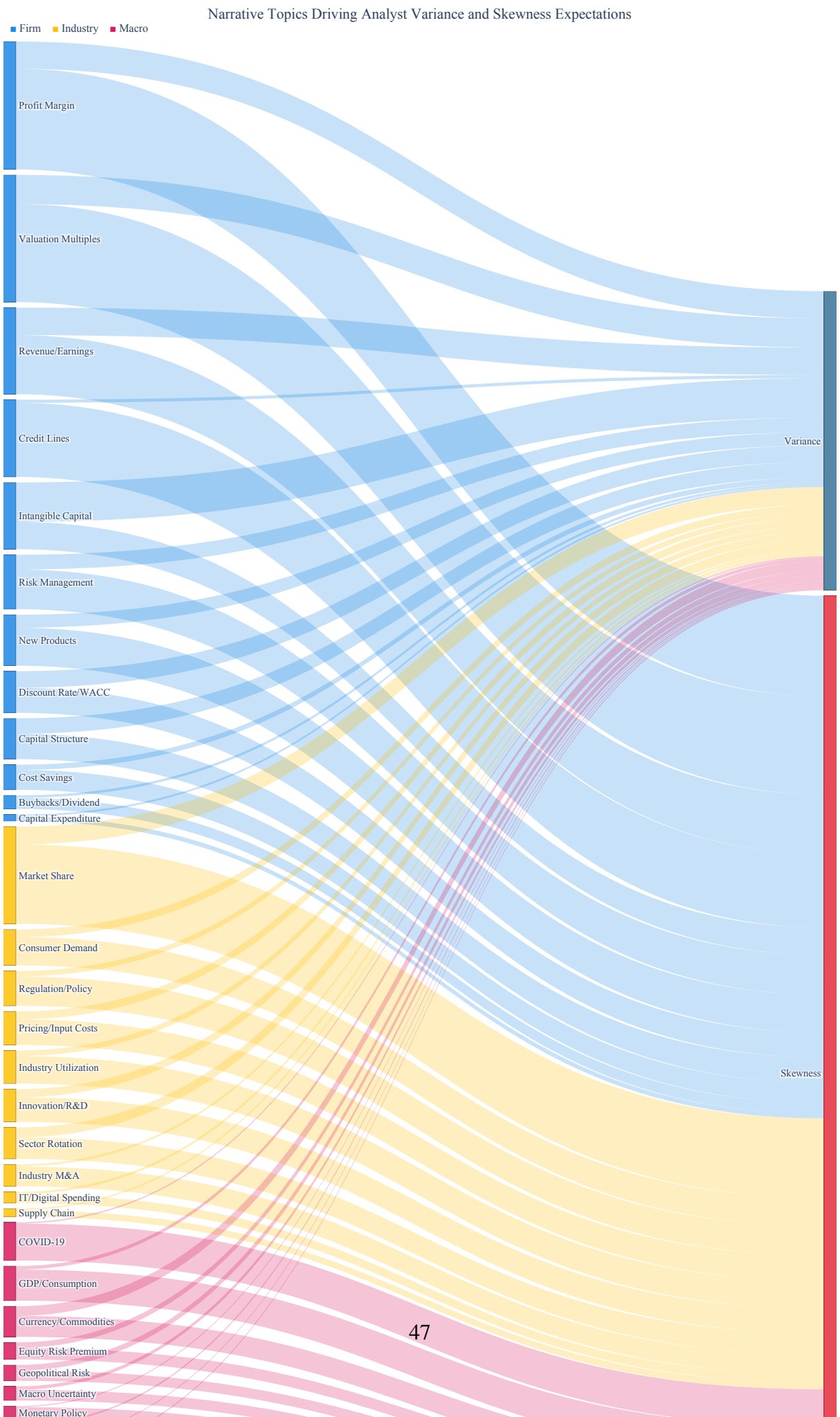

Narrative Topics Driving Analyst Variance and Skewness Expectations
Firm
Industry
Macro
Profit Margin
Valuation Multiples
Revenue/Earnings
Credit Lines
Intangible Capital
Risk Management
New Products
Discount Rate/WACC
Capital Structure
Cost Savings
Buybacks/Dividend
Capital Expenditure
Market Share
Consumer Demand
Regulation/Policy
Pricing/Input Costs
Industry Utilization
Innovation/R&D
Sector Rotation
Industry M&A
IT/Digital Spending
Supply Chain
COVID-19
GDP/Consumption
Currency/Commodities
Equity Risk Premium
Geopolitical Risk
Macro Uncertainty
Monetary Policy
Variance
Skewness

Table 1: **Summary Statistics**

Panel A reports summary statistics at the analyst report level. We construct subjective variance and skewness expectations from price-range forecasts extracted from analyst equity research reports issued by the largest 10 brokerage firms. Using a binomial tree model, we set the base case price equal to the analyst price target and define the probability $p = (\text{base} - \text{bear})/(\text{bull} - \text{bear})$ and log returns $r_u = \ln(\text{bull}/P)$ and $r_d = \ln(\text{bear}/P)$, where P is the current stock price when the report was issued. IV^{Option} and $Skew^{Option}$ are the model-free implied variance and skewness of Bakshi, Kapadia, and Madan (2003), computed from the volatility surface at a 365-day horizon. RV^{Past} and $Skew^{Past}$ are the annualized realized variance and skewness computed from daily log returns over the 252 trading days prior to each report date. The analyst's expected return is $r^{Analyst}_{i,t\to t+12} = \ln(\text{Price Target}/P_t)$, where P_t is the stock price on the report date. Firm characteristics include market beta (β), log market equity ($\log(ME)$), log book-to-market ($\log(B/M)$), asset growth (AG), and gross profitability (GPA), all measured as of the month-end prior to the report date; we require non-missing values for all characteristics in the analyst report sample. Panel B reports summary statistics of subjective variance and skewness by brokerage firm, ordered by sample size, with the start date indicating when each broker enters the sample. Panels C and D report summary statistics at the firm-month level for the straddle and skewness return samples. Each firm-month observation is matched to the most recent analyst forecast issued by each broker firm within the preceding 12 months. The consensus requires at least two contributing brokers and averages across them, and the contributing brokers' release dates are no more than six months apart. We compute both a simple-average and a recency-weighted consensus, where recency weights are $1/(\text{forecast lag} + 1)$. Straddle Return is the monthly at-the-money straddle return following Heston, Jones, Khorram, Li, and Mo (2023). The skewness asset return follows Bali and Murray (2013). Control variables include $\log(ME)$, IVOL (idiosyncratic volatility), Vol Deviation ($IV - RV$), IV Term Spread, IV Smirk, and VoV (volatility of volatility). The sample period is 2007 to 2024.

Panel A: Analyst Subjective Expectations (Report Level)

	N	Mean	Std	p5	p25	Median	p75	p95
$IV^{Analyst}_{t+1,t+12}$	180,858	0.187	0.339	0.015	0.044	0.087	0.187	0.695
$IV^{Option}_{t+1,t+12}$	180,858	0.269	0.373	0.048	0.086	0.146	0.290	0.894
$RV^{Past}_{t-12,t}$	180,858	0.202	0.275	0.030	0.061	0.114	0.235	0.657
$Skew^{Analyst}_{t+1,t+12}$	180,858	-0.357	0.744	-1.500	-0.765	-0.338	0.000	0.742
$Skew^{Option}_{t+1,t+12}$	180,858	-0.544	1.432	-1.160	-0.778	-0.571	-0.364	0.100
$Skew^{Past}_{t-12,t}$	180,858	-0.292	1.493	-2.654	-0.664	-0.161	0.275	1.525
$r^{Analyst}_{i,t\to t+12}$	180,858	0.143	0.230	-0.134	0.037	0.123	0.218	0.492
β	180,858	1.140	0.468	0.447	0.837	1.095	1.394	1.973
$\log(ME)$	180,858	9.17	1.74	6.35	7.98	9.12	10.34	12.15
$\log(B/M)$	180,858	-1.243	1.095	-3.158	-1.846	-1.128	-0.481	0.251
AG	180,858	0.286	11.235	-0.138	-0.005	0.059	0.174	0.817
GPA	180,858	0.267	0.246	0.000	0.117	0.243	0.395	0.688

Table 1: **Summary Statistics** (continued)

Panel B: Summary Statistics by Broker (Report Level)

			$IV^{Analyst}_{t+1,t+12}$			$Skew^{Analyst}_{t+1,t+12}$		
	Start	N	Mean	Median	Std	Mean	Median	Std
Barclays	2011/04	43,639	0.120	0.063	0.206	-0.469	-0.456	0.784
Morgan Stanley	2007/06	32,239	0.268	0.150	0.400	-0.107	-0.134	0.596
RBC	2008/01	31,396	0.147	0.075	0.267	-0.525	-0.516	0.706
Jefferies	2011/03	22,745	0.268	0.115	0.477	-0.265	-0.289	0.791
Credit Suisse	2010/01	17,766	0.188	0.092	0.341	-0.353	-0.338	0.801
UBS	2010/12	12,125	0.191	0.087	0.327	-0.324	-0.277	0.630
BMO	2016/08	11,858	0.199	0.066	0.380	-0.498	-0.474	0.780
Wells Fargo	2014/07	6,103	0.156	0.063	0.283	-0.245	-0.224	0.702
Evercore	2015/08	1,614	0.119	0.062	0.259	-0.293	-0.289	0.706
BofA	2011/07	1,373	0.071	0.014	0.181	-0.107	0.000	0.701

Panel C: Straddle Return Sample (Firm-Month Level)

	N	Mean	Std	p5	p25	Median	p75	p95
$IV^{Analyst}_{t+1,t+12}$ (Simple)	108,905	0.164	0.222	0.025	0.056	0.097	0.180	0.531
$IV^{Analyst}_{t+1,t+12}$ (Recency)	108,905	0.165	0.236	0.023	0.054	0.095	0.181	0.537
Straddle Return	108,905	-0.005	0.884	-0.928	-0.612	-0.192	0.377	1.447
log(ME)	108,905	23.01	1.50	20.63	21.97	22.96	24.00	25.62
IVOL	108,905	0.261	0.198	0.091	0.144	0.209	0.315	0.594
Vol Deviation	108,905	0.035	0.268	-0.380	-0.129	0.043	0.198	0.433
IV Term Spread	108,905	-0.003	0.069	-0.083	-0.019	0.003	0.020	0.063
IV Smirk	108,905	0.006	0.155	-0.084	-0.034	-0.018	0.002	0.192
VoV	108,898	0.075	0.073	0.028	0.041	0.056	0.080	0.191

Panel D: Skewness Return Sample (Firm-Month Level)

	N	Mean	Std	p5	p25	Median	p75	p95
$Skew^{Analyst}_{t+1,t+12}$ (Simple)	103,769	-0.361	0.466	-1.096	-0.645	-0.369	-0.084	0.392
$Skew^{Analyst}_{t+1,t+12}$ (Recency)	103,769	-0.369	0.525	-1.196	-0.679	-0.366	-0.057	0.459
Skewness Return	103,769	-0.006	0.148	-0.146	-0.047	-0.002	0.046	0.136
log(ME)	103,769	23.08	1.48	20.75	22.04	23.03	24.06	25.66
IVOL	103,769	0.260	0.195	0.091	0.145	0.209	0.314	0.591
Vol Deviation	102,675	0.035	0.262	-0.376	-0.128	0.045	0.199	0.431
IV Term Spread	103,769	-0.003	0.063	-0.080	-0.019	0.003	0.020	0.061
IV Smirk	103,769	0.002	0.132	-0.076	-0.033	-0.018	-0.000	0.149
VoV	103,764	0.070	0.062	0.028	0.041	0.055	0.077	0.165

Table 2: **Predicting Realized Variance and Skewness**

This table reports results from regressions of future realized variance and skewness on analyst subjective expectations, option-implied moments, and past realized moments. The dependent variable in Panel A is annualized realized variance ($RV^{Realized}_{t+1,t+12}$), computed from daily log returns over the 252 trading days following each analyst report date; in Panel B, it is realized skewness ($Skew^{Realized}_{t+1,t+12}$), the standardized third central moment over the same window. $IV^{Analyst}$ and $Skew^{Analyst}$ are the subjective variance and skewness expectations, respectively, derived from price-range forecasts in analyst equity research reports. IV^{Option} and $Skew^{Option}$ are the model-free option-implied variance and skewness of Bakshi, Kapadia, and Madan (2003), computed from the volatility surface at a 365-day horizon. RV^{Past} and $Skew^{Past}$ are the realized variance and skewness over the prior 252 trading days. Firm characteristics (market beta, log market equity, log book-to-market, asset growth, and gross profitability-to-asset) are included as controls where indicated. Specifications progressively add broker and year-month fixed effects. All independent variables are winsorized at the 1st and 99th percentiles within each year-month. Standard errors are three-way clustered by broker, firm, and year-month; t-statistics are reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is 2007 to 2024.

Panel A: Predicting Future Realized Variance

	$RV^{Realized}_{t+1,t+12}$			
	(1)	(2)	(3)	(4)
$IV^{Analyst}_{t+1,t+12}$	0.566***	0.180***	0.157***	0.159***
	[12.99]	[8.11]	[8.28]	[8.33]
$IV^{Option}_{t+1,t+12}$		0.350***	0.333***	0.331***
		[12.13]	[10.87]	[10.94]
$RV^{Past}_{t-12,t}$		0.158**	0.305***	0.306***
		[2.93]	[6.67]	[6.77]
Firm Characteristics	No	Yes	Yes	Yes
Broker FE	No	No	Yes	No
Year-Month FE	No	No	Yes	No
Broker×Year-Month FE	No	No	No	Yes
N	180,206	180,206	180,205	180,143
Adj. R^2	0.181	0.312	0.376	0.382

Panel B: Predicting Future Realized Skewness

	$Skew^{Realized}_{t+1,t+12}$			
	(1)	(2)	(3)	(4)
$Skew^{Analyst}_{t+1,t+12}$	0.089***	0.066***	0.074***	0.075***
	[8.48]	[4.82]	[4.18]	[4.12]
$Skew^{Option}_{t+1,t+12}$		0.014	0.021	0.022
		[0.32]	[0.51]	[0.53]
$Skew^{Past}_{t-12,t}$		0.003	0.002	0.002
		[0.33]	[0.19]	[0.24]
Firm Characteristics	No	Yes	Yes	Yes
Broker FE	No	No	Yes	No
Year-Month FE	No	No	Yes	No
Broker×Year-Month FE	No	No	No	Yes
N	180,206	180,206	180,205	180,143
Adj. R^2	0.001	0.004	0.013	0.020

Table 3: **Broker-Level Forecasting Regressions**

This table reports results from estimating the baseline specification in Table 3 separately within each broker's coverage universe. In Panel A, the dependent variable is annualized realized variance ($RV^{Realized}_{t+1,t+12}$); in Panel B, it is realized skewness ($Skew^{Realized}_{t+1,t+12}$). The key regressors are the analyst's subjective expectation ($IV^{Analyst}$ or $Skew^{Analyst}$), the option-implied moment (IV^{Option} or $Skew^{Option}$), and the past realized moment (RV^{Past} or $Skew^{Past}$). The same set of firm characteristics in Table 2 is included as controls. All specifications absorb year-month fixed effects. Independent variables are winsorized at the 1st and 99th percentiles. Standard errors are two-way clustered by firm and year-month; t-statistics appear in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel A: Predicting Realized Variance

	$RV^{Realized}_{t+1,t+12}$									
	(1) Barclays	(2) Morgan Stanley	(3) RBC	(4) Jefferies	(5) Credit Suisse	(6) UBS	(7) BMO	(8) Wells Fargo	(9) Evercore	(10) BofA
$IV^{Analyst}_{t+1,t+12}$	0.195***	0.154***	0.260***	0.119***	0.191***	0.075**	0.123***	0.091	0.232**	0.355**
	[4.20]	[6.58]	[5.52]	[3.38]	[2.89]	[2.38]	[4.01]	[1.14]	[2.11]	[2.00]
$IV^{Option}_{t+1,t+12}$	0.326***	0.340***	0.346***	0.429***	0.472***	0.517***	0.323***	0.299***	0.334***	0.609***
	[7.27]	[7.79]	[8.55]	[9.54]	[4.32]	[5.97]	[8.10]	[4.79]	[2.69]	[2.72]
$RV^{Past}_{t-12,t}$	0.349***	0.345***	0.242***	0.279***	0.359***	0.311***	0.269***	0.542***	0.180	0.063
	[6.35]	[6.59]	[4.00]	[4.78]	[2.74]	[3.89]	[5.08]	[3.19]	[1.35]	[0.29]
Firm Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	43,507	32,143	31,234	22,639	17,650	12,100	11,829	6,083	1,602	1,356
Adj. R^2	0.373	0.426	0.440	0.338	0.372	0.255	0.545	0.498	0.513	0.507

Panel B: Predicting Realized Skewness

	$Skew^{Realized}_{t+1,t+12}$									
	(1) Barclays	(2) Morgan Stanley	(3) RBC	(4) Jefferies	(5) Credit Suisse	(6) UBS	(7) BMO	(8) Wells Fargo	(9) Evercore	(10) BofA
$Skew^{Analyst}_{t+1,t+12}$	0.124***	0.072**	0.050	0.026	0.048	0.139***	0.057	0.029	0.076	0.051
	[4.57]	[2.13]	[1.51]	[0.83]	[1.35]	[3.12]	[1.20]	[0.54]	[0.78]	[0.61]
$Skew^{Option}_{t+1,t+12}$	0.014	0.072	0.002	-0.077	0.104	0.159*	0.076	0.069	-0.374*	-0.122
	[0.23]	[1.20]	[0.03]	[-1.03]	[0.93]	[1.89]	[0.69]	[0.47]	[-1.87]	[-0.70]
$Skew^{Past}_{t-12,t}$	-0.008	0.011	0.001	0.008	-0.003	-0.019	-0.013	0.014	0.017	0.041
	[-0.42]	[0.66]	[0.08]	[0.49]	[-0.13]	[-0.94]	[-0.44]	[0.40]	[0.22]	[0.58]
Firm Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	43,507	32,143	31,234	22,639	17,650	12,100	11,829	6,083	1,602	1,356
Adj. R^2	0.020	0.018	0.018	0.015	0.025	0.028	0.018	0.028	0.108	0.179

Table 4: **Predicting Option Returns: Fama–MacBeth Regressions**

This table reports monthly Fama and MacBeth (1973) cross-sectional regression results for predicting option returns using analyst subjective expectations. The dependent variable in Panel A is the monthly straddle return (%), constructed from at-the-money calls and puts formed on the third Friday of each month and held to expiration; in Panel B, it is the monthly skewness asset return (%) of Bali and Murray (2013). Column (1) uses the equal-weighted consensus measure, defined as the simple average of each analyst's subjective expectation across all brokers covering the same firm in a given month. Column (2) uses the recency-weighted consensus, which weights each broker's forecast inversely by its staleness. Columns (3)–(5) report results estimated separately for the three largest brokers with sufficiently large firm coverage over a long sample period: Barclays, Morgan Stanley, and RBC. The key regressor is the analyst's subjective expectation ($IV^{Analyst}_{t+1,t+12}$ or $Skew^{Analyst}_{t+1,t+12}$). Controls include the option-related predictors described in Section 2.4: log market capitalization, idiosyncratic volatility, volatility deviation, IV term spread, IV smirk slope, volatility of volatility, the lagged option return (short-term reversal), and 12-2 month average option return (momentum). All independent variables are winsorized at the 1st and 99th percentiles and then cross-sectionally standardized to zero mean and unit variance within each month, so that coefficients represent the effect of a one-standard-deviation increase. Cross-sections with fewer than 100 firms with valid analyst expectations are excluded. Standard errors are computed using Newey–West with 12 lags; t-statistics appear in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel A: Predicting Straddle Returns

	Straddle Return (%)				
	(1) EW Consensus	(2) RW Consensus	(3) Morgan Stanley	(4) Barclays	(5) RBC
$IV^{Analyst}_{t+1,t+12}$	1.145***	1.043***	1.897***	1.647***	0.679*
	[2.69]	[2.66]	[3.53]	[4.29]	[1.89]
log(Market Cap)	0.966**	0.926**	0.976*	0.688	0.941
	[2.12]	[2.12]	[1.85]	[1.28]	[1.62]
Idiosyncratic Volatility	-1.916***	-1.896***	-2.468***	-1.624**	-1.407**
	[-3.35]	[-3.24]	[-4.31]	[-2.28]	[-2.20]
Volatility Deviation	0.911**	0.948**	1.229**	0.929*	0.817
	[2.19]	[2.28]	[2.09]	[1.91]	[1.65]
IV Term Spread	0.488	0.478	0.463	0.081	0.597
	[0.78]	[0.76]	[0.79]	[0.16]	[0.88]
IV Smirk Slope	0.125	0.128	0.167	0.163	-0.186
	[0.44]	[0.46]	[0.61]	[0.44]	[-0.60]
Volatility of Volatility	-1.437***	-1.457***	-1.270***	-1.559***	-1.481***
	[-4.01]	[-4.07]	[-4.35]	[-3.60]	[-4.30]
Lagged Straddle Return	-0.801*	-0.804*	-0.725**	-0.841	-0.823*
	[-1.85]	[-1.85]	[-2.17]	[-1.51]	[-1.68]
Straddle Momentum	1.016***	1.038***	0.794*	0.968**	1.203***
	[2.86]	[2.91]	[1.94]	[2.52]	[2.62]
N	93,983	93,983	67,897	75,474	59,628
Adj. R^2	0.022	0.022	0.022	0.024	0.025

Panel B: Predicting Skewness Asset Returns

	Skewness Asset Return (%)				
	(1) EW Consensus	(2) RW Consensus	(3) Morgan Stanley	(4) Barclays	(5) RBC
$Skew^{Analyst}_{t+1,t+12}$	0.125***	0.141**	0.233**	0.052	0.096**
	[2.80]	[2.47]	[2.59]	[0.88]	[2.24]
log(Market Cap)	0.071	0.071	0.134	0.075	0.102
	[0.96]	[0.94]	[1.39]	[0.67]	[1.14]
Idiosyncratic Volatility	0.017	0.014	0.095	0.016	0.005
	[0.24]	[0.21]	[1.42]	[0.22]	[0.05]
Volatility Deviation	-0.069	-0.067	-0.072	0.086	-0.065
	[-0.83]	[-0.81]	[-0.94]	[1.62]	[-0.51]
IV Term Spread	0.054	0.055	0.095	-0.050	-0.048
	[0.72]	[0.73]	[0.96]	[-0.72]	[-0.57]
IV Smirk Slope	-0.200***	-0.201***	-0.280***	-0.130	-0.166**
	[-3.21]	[-3.29]	[-3.35]	[-1.40]	[-2.21]
Volatility of Volatility	0.100*	0.101*	0.120**	0.125*	0.115*
	[1.84]	[1.89]	[2.49]	[1.95]	[1.77]
Lagged Skewness Return	-0.012	-0.012	-0.059	0.068	-0.005
	[-0.20]	[-0.20]	[-0.85]	[1.18]	[-0.10]
Skewness Momentum	0.044	0.042	0.127	0.143**	0.045
	[0.72]	[0.68]	[1.35]	[2.01]	[0.74]
N	86,456	86,456	62,421	69,183	54,536
Adj. R^2	0.030	0.030	0.033	0.031	0.035

Table 5: **Predicting Option Returns: Portfolio Sorts**

This table reports the average monthly return, t-statistic, and annualized Sharpe ratio of quintile long-short (H−L) portfolios sorted on orthogonalized analyst expectations. Each month, the analyst signal is orthogonalized by cross-sectionally regressing it on the full set of controls in Table 4 and retaining the residual, so that the portfolio spread isolates the incremental information in analyst expectations beyond known option return predictors. Firms are then sorted into quintiles based on the orthogonalized signal, and the long-short portfolio buys the highest quintile and sells the lowest. Panel A uses the equal-weighted consensus signal; Panels B–D use individual broker signals for the three largest brokers: Barclays, Morgan Stanley, and RBC. Results are reported for the full sample as well as the S&P 500 and Russell 1000 universes, where options are highly liquid. EW = equally weighted; VW = value-weighted. t-statistics are computed using Newey–West with 12 lags. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Full Sample		**S&P 500**		**Russell 1000**	
	EW	VW	EW	VW	EW	VW
Panel A: Consensus						
Straddle H−L						
Return (%)	4.61***	7.98***	4.86***	5.65***	5.37***	6.87***
t-statistic	[6.02]	[4.70]	[3.96]	[2.93]	[5.92]	[3.67]
Sharpe ratio	1.22	1.04	0.86	0.65	1.21	0.87
Skewness H−L						
Return (%)	0.38***	0.31*	0.36**	0.18	0.41***	0.28
t-statistic	[3.38]	[1.86]	[2.23]	[1.01]	[3.30]	[1.59]
Sharpe ratio	0.64	0.47	0.57	0.25	0.64	0.41
Panel B: Barclays						
Straddle H−L						
Return (%)	5.67***	9.92***	7.15***	5.03***	7.41***	7.28**
t-statistic	[3.80]	[3.67]	[4.37]	[3.14]	[3.97]	[2.19]
Sharpe ratio	1.13	0.94	1.16	0.62	1.22	0.65
Skewness H−L						
Return (%)	0.17	0.29	0.04	0.29	0.04	0.32
t-statistic	[1.20]	[1.54]	[0.35]	[1.55]	[0.26]	[1.48]
Sharpe ratio	0.22	0.32	0.07	0.29	0.05	0.34
Panel C: Morgan Stanley						
Straddle H−L						
Return (%)	5.30***	7.91***	5.74***	8.43***	5.45***	6.72***
t-statistic	[4.11]	[3.45]	[4.29]	[4.56]	[4.73]	[3.22]
Sharpe ratio	1.24	0.87	1.08	0.94	1.26	0.78
Skewness H−L						
Return (%)	0.64**	0.97***	0.69**	1.00***	0.53**	0.90***
t-statistic	[2.48]	[3.12]	[2.11]	[3.02]	[2.02]	[3.27]
Sharpe ratio	0.65	0.68	0.59	0.72	0.56	0.70
Panel D: RBC						
Straddle H−L						
Return (%)	3.29***	7.40***	4.29**	4.66	4.80***	7.40***
t-statistic	[2.94]	[2.70]	[2.21]	[1.56]	[3.46]	[3.25]
Sharpe ratio	0.75	0.90	0.82	0.59	1.04	0.93
Skewness H−L						
Return (%)	0.14	0.02	0.34*	-0.05	0.22	-0.04
t-statistic	[0.95]	[0.05]	[1.73]	[-0.14]	[1.16]	[-0.13]
Sharpe ratio	0.21	0.02	0.54	-0.05	0.30	-0.04

Table 6: **Variance Decomposition: Valuation Multiples vs. Earnings**

This table decomposes the predictive content of analyst subjective expectations into a multiple (valuation) component and an EPS (earnings) component. Since $\log(\text{PT}/P) = \log(\text{Multiple}) + \log(\text{EPS}/P)$, variance decomposes additively in logs. $IV^{Multiple}$ ($Skew^{Multiple}$) captures uncertainty from the bull and bear valuation multiples, and IV^{EPS} ($Skew^{EPS}$) captures uncertainty from the bull and bear EPS forecasts. Both use the same binomial probability $p = (\text{base} - \text{bear})/(\text{bull} - \text{bear})$ derived from the price targets. Columns (1) and (4) report the baseline specification using the aggregate analyst expectation; columns (2) and (5) replace it with the multiple component; columns (3) and (6) replace it with the EPS component. Each component is estimated on its own available sample. All specifications include broker and year-month fixed effects, firm characteristics (market beta, log market equity, log book-to-market, asset growth, and gross profitability), with independent variables winsorized at the 1st and 99th percentiles within each year-month. Standard errors are two-way clustered by firm and year-month; t-statistics are reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$RV^{Realized}_{t+1,t+12}$			$Skew^{Realized}_{t+1,t+12}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$IV^{Analyst}_{t+1,t+12}$	0.116***					
	[5.61]					
$IV^{Multiple}$		0.027**				
		[2.37]				
IV^{EPS}			0.016			
			[1.45]			
$IV^{Option}_{t+1,t+12}$	0.293***	0.303***	0.291***			
	[7.20]	[7.15]	[4.45]			
$RV^{Past}_{t-12,t}$	0.437***	0.467***	0.345***			
	[6.12]	[6.33]	[4.02]			
$Skew^{Analyst}_{t+1,t+12}$				0.068***		
				[4.23]		
$Skew^{Multiple}$					0.052***	
					[3.25]	
$Skew^{EPS}$						0.047**
						[2.14]
$Skew^{Option}_{t+1,t+12}$				-0.009	-0.009	-0.073
				[-0.21]	[-0.19]	[-1.27]
$Skew^{Past}_{t-12,t}$				-0.000	0.000	-0.007
				[-0.01]	[0.01]	[-0.34]
Firm Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Broker FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	63,806	60,788	24,490	63,806	60,788	24,490
Adj. R^2	0.342	0.334	0.277	0.015	0.015	0.021

Table 7: **Hedge Fund Derivative Ownership and Analyst Expectations**

This table examines whether hedge fund (alternative asset manager) derivative positions respond to analyst subjective expectations for variance and skewness. The dependent variables are hedge fund call ownership and put ownership, expressed as a percentage of shares outstanding. Hedge fund option holdings data sourced from SEC 13F forms (see, Aragon, Chen, and Shi, 2022). Columns (1)–(2) include analyst subjective variance ($IV^{Analyst}$), option-implied variance (IV^{Option}), and past realized variance (RV^{Past}). Columns (3)–(4) include analyst subjective skewness ($Skew^{Analyst}$), option-implied skewness ($Skew^{Option}$), and past realized skewness ($Skew^{Past}$). Columns (5)–(6) include both variance and skewness measures simultaneously. All specifications include firm and year-month fixed effects, option-level controls (log market capitalization, idiosyncratic volatility, volatility deviation, implied volatility term spread, implied volatility smirk, and volatility of volatility), and firm characteristics (market beta, log book-to-market, asset growth, and gross profitability-to-asset). All independent variables are winsorized at the 1st and 99th percentiles within each year-month. Standard errors are two-way clustered by firm and year-month; t-statistics are reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is 2011 to 2024.

	Variance		Skewness		Variance & Skewness	
	(1) Call	(2) Put	(3) Call	(4) Put	(5) Call	(6) Put
$IV^{Analyst}_{t+1,t+12}$	0.156**	0.159**			0.139*	0.140*
	[2.18]	[2.00]			[1.97]	[1.76]
$Skew^{Analyst}_{t+1,t+12}$			0.029**	0.037***	0.022**	0.031**
			[2.49]	[3.01]	[1.99]	[2.53]
$IV^{Option}_{t+1,t+12}$	0.073	0.153**			0.042	0.131**
	[1.38]	[2.47]			[0.79]	[2.12]
$Skew^{Option}_{t+1,t+12}$			0.182***	0.134***	0.124***	0.082***
			[8.16]	[4.75]	[6.02]	[3.02]
$RV^{Past}_{t-12,t}$	0.479***	0.290***			0.437***	0.263***
	[5.79]	[3.95]			[5.32]	[3.52]
$Skew^{Past}_{t-12,t}$			-0.007*	-0.004	-0.003	-0.001
			[-1.78]	[-0.79]	[-0.82]	[-0.29]
Option Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	57,102	57,102	57,102	57,102	57,102	57,102
Adj. R^2	0.582	0.641	0.578	0.639	0.583	0.642

Table 8: **Risk–Return and Skewness–Return Trade-Offs**

This table reports panel regressions of the analyst's expected stock return on analyst subjective expectations and firm characteristics at the broker–firm–date level. The dependent variable is the analyst's log return expectation, $r^{Analyst}_{b,i,t\to t+12} = \ln(\text{Price Target}/P_t)$, where b indexes brokers, i indexes firms, and date t is the current stock price date reported in each analyst report. In Panel A, the key regressor is the analyst's subjective variance ($IV^{Analyst}_{t+1,t+12}$); in Panel B, it is the analyst's subjective skewness ($Skew^{Analyst}_{t+1,t+12}$). Columns (1)–(2) use all analyst reports with broker and year-month fixed effects; column (1) includes only the analyst expectation, and column (2) adds firm characteristics: market beta, log market capitalization (log(ME)), log book-to-market (log(B/M)), asset growth (AG), and gross profitability-to-asset (GPA). Columns (3)–(5) report results estimated separately for the three largest brokers with sufficiently large firm coverage over a long sample period: Barclays, Morgan Stanley, and RBC, each with year-month fixed effects and firm characteristics. Standard errors are three-way clustered by broker, firm, and year-month for columns (1)–(2), and two-way clustered by firm and year-month for columns (3)–(5); t-statistics appear in brackets. All independent variables are winsorized at the 1st and 99th percentiles. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel A: Risk-Return Trade-Off

	$r^{Analyst}_{i,t\to t+12}$				
	(1) All	(2) All	(3) Barclays	(4) Morgan Stanley	(5) RBC
$IV^{Analyst}_{t+1,t+12}$	0.258***	0.203***	0.152*	0.067***	0.372***
	[4.40]	[3.83]	[1.79]	[2.69]	[19.86]
β		0.026***	0.028***	0.032***	0.039***
		[3.76]	[3.32]	[3.33]	[5.51]
$\log(ME)$		-0.005	-0.002	0.007**	-0.007***
		[-1.37]	[-0.53]	[2.25]	[-3.59]
$\log(B/M)$		-0.005	0.002	0.006	-0.000
		[-1.37]	[0.53]	[1.37]	[-0.16]
GPA		-0.119***	-0.125***	-0.090***	-0.043***
		[-8.27]	[-6.64]	[-3.86]	[-2.79]
AG		0.032***	0.032***	0.035***	0.035***
		[10.68]	[4.96]	[4.86]	[6.51]
Broker FE	Yes	Yes	No	No	No
Year-Month FE	Yes	Yes	Yes	Yes	Yes
N	180,857	180,857	43,639	32,237	31,395
Adj. R^2	0.186	0.211	0.132	0.108	0.355

Panel B: Skewness-Return Trade-Off

	$r^{Analyst}_{i,t\to t+12}$				
	(1) All	(2) All	(3) Barclays	(4) Morgan Stanley	(5) RBC
$Skew^{Analyst}_{t+1,t+12}$	-0.076***	-0.089***	-0.090***	-0.135***	-0.062***
	[-8.42]	[-11.24]	[-18.67]	[-20.10]	[-13.69]
β		0.053***	0.049***	0.056***	0.061***
		[8.00]	[7.29]	[6.51]	[7.61]
$\log(ME)$		-0.022***	-0.014***	-0.007**	-0.028***
		[-4.07]	[-5.07]	[-2.21]	[-9.45]
$\log(B/M)$		-0.007	0.002	0.004	-0.008**
		[-1.74]	[0.61]	[1.26]	[-2.16]
GPA		-0.160***	-0.118***	-0.098***	-0.139***
		[-6.43]	[-5.77]	[-4.97]	[-6.50]
AG		0.036***	0.032***	0.039***	0.039***
		[10.56]	[5.35]	[5.59]	[6.47]
Broker FE	Yes	Yes	No	No	No
Year-Month FE	Yes	Yes	Yes	Yes	Yes
N	180,857	180,857	43,639	32,237	31,395
Adj. R^2	0.137	0.241	0.234	0.231	0.256

Table 9: **Determinants of Analyst Subjective Variance and Skewness**

This table reports panel regressions of analyst subjective expectations on option-implied moments, past realized moments, and firm characteristics at the broker–firm–date level. In Panel A, the dependent variable is the analyst's subjective variance ($IV^{Analyst}_{t+1,t+12}$); in Panel B, it is the analyst's subjective skewness ($Skew^{Analyst}_{t+1,t+12}$). Column (1) includes the option-implied and past realized moments without fixed effects; column (2) adds firm characteristics: market beta, log market capitalization (log(ME)), log book-to-market (log(B/M)), asset growth (AG), and gross profitability-to-asset (GPA), without fixed effects; column (3) further adds broker and year-month fixed effects. Standard errors are three-way clustered by broker, firm, and year-month; t-statistics appear in brackets. All independent variables are winsorized at the 1st and 99th percentiles. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel A: Determinants of Analyst Subjective Variance

	$IV^{Analyst}_{t+1,t+12}$		
	(1)	(2)	(3)
$IV^{Option}_{t+1,t+12}$	0.434***	0.404***	0.386***
	[10.15]	[12.64]	[11.31]
$RV^{Past}_{t-12,t}$	0.303***	0.270***	0.328***
	[4.33]	[4.36]	[4.89]
β		0.011	-0.005
		[0.66]	[-0.35]
$\log(ME)$		-0.006	-0.005
		[-1.44]	[-1.31]
$\log(B/M)$		-0.019**	-0.015**
		[-3.06]	[-2.45]
GPA		-0.170***	-0.183***
		[-3.43]	[-3.75]
AG		0.001	-0.001
		[0.09]	[-0.26]
Broker FE	No	No	Yes
Year-Month FE	No	No	Yes
N	180,858	180,858	180,857
Adj. R^2	0.326	0.337	0.365

Panel B: Determinants of Analyst Subjective Skewness

	$Skew^{Analyst}_{t+1,t+12}$		
	(1)	(2)	(3)
$Skew^{Option}_{t+1,t+12}$	0.342***	0.262***	0.242***
	[10.63]	[11.03]	[14.60]
$RSkew^{Past}_{t-12,t}$	-0.052***	-0.044***	-0.041***
	[-13.10]	[-12.68]	[-11.22]
β		0.023	0.012
		[0.81]	[0.57]
$\log(ME)$		-0.061***	-0.064***
		[-8.95]	[-9.90]
$\log(B/M)$		0.033	0.038**
		[1.70]	[2.72]
GPA		0.155**	0.134*
		[3.21]	[2.05]
AG		-0.054**	-0.049***
		[-3.20]	[-3.49]
Broker FE	No	No	Yes
Year-Month FE	No	No	Yes
N	180,858	180,858	180,857
Adj. R^2	0.043	0.067	0.113

Table 10: **Topic Composition of Analyst Information Sets**

This table reports the average topic composition of analyst scenario discussions for Barclays, Morgan Stanley, and RBC. Panel A reports entity-level weights: the fraction of the scenario text attributed to firm-specific, industry, or macroeconomic topics. Panel B reports the conditional probability of each sub-topic within its entity category, summing to 100% within each of the three entity groups. Topics are classified by a large language model (Gemini 2.0 Flash) from the original scenario text. Columns show the overall average and the averages for base, bull, and bear scenarios separately. The last column reports the difference between bear and bull scenarios.

	All	Base	Bull	Bear	Bear−Bull
Panel A: Entity-Level Weights (%)					
Firm	76.8	81.1	79.4	69.9	-9.5
Industry	16.9	15.8	14.5	20.3	+5.8
Macroeconomy	6.3	3.1	6.1	9.7	+3.6
Panel B: Sub-Topic Weights (%, Conditional Probability)					
Firm					
Revenue/Earnings Growth	27.2	26.8	30.4	24.5	-5.9
Valuation Multiples	25.6	26.2	25.3	25.3	-0.1
Risk Management	15.7	15.6	9.5	22.0	+12.5
Profit Margin	10.3	9.6	11.0	10.5	-0.5
New Products/Launches	5.4	5.1	7.7	3.5	-4.2
Intangible Capital	3.7	3.3	3.9	4.0	+0.2
Capital Structure	3.0	3.5	3.2	2.2	-1.0
Cost Savings/Op. Leverage	2.8	2.8	3.5	2.0	-1.5
Credit Lines	2.6	2.1	2.4	3.3	+0.8
Discount Rate/WACC	1.7	2.5	1.2	1.2	+0.0
Buybacks/Dividends	1.5	2.1	1.4	0.9	-0.5
Capital Expenditure	0.5	0.4	0.4	0.6	+0.2
Industry					
Regulation/Policy	20.6	20.9	18.8	22.1	+3.3
Market Share/Competition	18.9	20.5	16.2	19.9	+3.7
Consumer Demand	11.7	10.4	12.9	11.7	-1.2
Pricing/Input Costs	10.9	8.6	10.0	14.0	+4.0
Innovation/R&D	9.0	10.2	9.7	7.2	-2.5
Sector Rotation/Cyclical	6.6	6.8	6.8	6.3	-0.5
Industry Utilization	6.3	6.1	7.4	5.5	-2.0
IT/Digital Spending	5.8	6.0	6.2	5.3	-0.9
Industry M&A	5.4	5.6	6.7	4.0	-2.7
Supply Chain	4.7	4.9	5.2	4.0	-1.2
Macroeconomy					
GDP/Consumer Spending	14.4	11.8	15.3	16.2	+0.9
Macro Uncertainty	12.2	11.6	10.2	14.9	+4.8
Currency/Commodities	10.7	10.2	11.1	10.9	-0.2
Geopolitical Risk	10.4	10.3	10.7	10.0	-0.7
Inflation/Yield Curve	9.4	9.7	9.4	9.2	-0.2
Equity Risk Premium	9.4	9.8	9.4	8.9	-0.5
Monetary Policy/Rates	9.0	9.7	9.3	8.0	-1.3
COVID-19 Impact	8.3	9.2	8.2	7.6	-0.7
Housing Conditions	8.1	8.8	8.2	7.3	-0.9
Fiscal Policy	8.0	8.8	8.2	7.0	-1.2
Observations	315,309	105,103	105,103	105,103	

Table 11: **Out-of-Sample R^2 by Information Set**

This table reports the out-of-sample R^2 from XGBoost models predicting analyst subjective variance and skewness. We randomly hold out 20% of observations (stratified by broker) as a test set. Hyperparameters (tree depth and number of estimators) are selected via 5-fold cross-validation on the remaining 80% training set, and each model is then trained on the full training set and evaluated on the held-out test set. Panel A uses market-based predictors (option-implied and past realized moments) as the baseline, with columns progressively adding narrative topic features. Panel B augments the baseline with firm characteristics (market beta, log market capitalization, log book-to-market, gross profitability, and asset growth) and evaluates the incremental contribution of narrative topics on top of this richer benchmark. Topic features are the conditional sub-topic weights of 32 sub-topics under bear and bull scenarios ($2 \times 32 = 64$ topics in total).

Panel A: Market Moments					
	Market Only	+ All Topics	+ Firm Topics	+ Ind. Topics	+ Macro Topics
No. of features	2	66	26	22	22
Analyst Variance	30.60%	53.98%	50.99%	35.68%	29.28%
Analyst Skewness	4.65%	29.25%	23.93%	11.68%	10.02%
Panel B: Market Moments + Firm Characteristics					
	Market + Chars	+ All Topics	+ Firm Topics	+ Ind. Topics	+ Macro Topics
No. of features	7	71	31	27	27
Analyst Variance	40.40%	59.89%	58.43%	50.85%	44.31%
Analyst Skewness	21.33%	38.74%	37.64%	29.95%	27.66%

Appendix

A.1. Additional Figures and Tables

Barclays | Coinbase Global, Inc.

U.S. Brokers, Asset Managers & Exchanges — POSITIVE

Coinbase Global, Inc. (COIN) — EQUAL WEIGHT

Income statement ($mn)	**2023A**	**2024E**	**2025E**	**2026E**	**CAGR**
Revenue	3,108	5,458	4,681	5,106	18.0%
EBITDA (adj)	965	2,473	1,853	2,242	32.5%
EBIT (adj)	825	2,337	1,713	2,102	36.6%
Pre-tax income (adj)	-77	1,689	804	1,192	N/A
Net income (adj)	590	1,805	1,305	1,616	39.9%
Margin and return data	**2023A**	**2024E**	**2025E**	**2026E**	**Average**
EBITDA (adj) margin (%)	31.0	45.3	39.6	43.9	40.0
EBIT (adj) margin (%)	26.5	42.8	36.6	41.2	36.8
Pre-tax (adj) margin (%)	-2.5	30.9	17.2	23.3	17.2
Net (adj) margin (%)	19.0	33.1	27.9	31.6	27.9
ROE (%)	10.0	25.5	19.5	24.7	19.9
Balance sheet ($mn)	**2023A**	**2024E**	**2025E**	**2026E**	**CAGR**
Cash and equivalents	5,139	7,167	7,296	7,499	13.4%
Total assets	206,983	397,239	480,005	580,749	41.0%
Short and long-term debt	2,980	4,220	4,220	4,220	12.3%
Total liabilities	200,701	390,757	473,316	573,547	41.9%
Net debt/(funds)	2,082	2,360	2,489	2,692	8.9%
Shareholders' equity	6,282	6,482	6,690	7,202	4.7%
Cash flow from operations	923	1,423	1,165	1,446	16.1%
Capital expenditure	0	0	0	0	N/A
Free cash flow	923	1,423	1,165	1,446	16.1%
Valuation metrics	**2023A**	**2024E**	**2025E**	**2026E**	**Average**
P/E (adj) (x)	70.0	25.6	37.2	31.6	41.1
Dividend yield (%)	0.0	0.0	0.0	0.0	0.0
Selected operating metrics ($mn)	**2023A**	**2024E**	**2025E**	**2026E**	
Transaction revenue	1,487	2,771	2,321	2,608	
Subscription & services revenue	1,407	2,234	1,998	2,114	
Net revenue	3,108	5,458	4,681	5,106	
Retail volume ($bn)	75	144	120	137	
Institutional volume ($bn)	392	769	743	833	
Retail transaction revenue	1,397	2,513	2,071	2,328	
Institutional transaction revenue	90	258	250	280	
Blockchain rewards revenue	331	645	652	706	
Custodial fee revenue	70	135	144	151	
Earn campaign revenue	0	0	0	0	
MTUs (mn)	7	8	8	9	
Per share data ($)	**2023A**	**2024E**	**2025E**	**2026E**	**CAGR**
EPS (adj)	2.44	6.69	4.60	5.41	30.4%
Diluted shares (mn)	242	270	284	299	7.3%
DPS	0.00	0.00	0.00	0.00	N/A
Payout ratio (%)	N/A	N/A	N/A	N/A	N/A

Note: FY End Dec
Source: Company data, Bloomberg, Barclays Research

Price (04-Oct-2024) **USD 170.91**
Price Target **USD 175.00**

Why EQUAL WEIGHT?

COIN is positioned as the dominant player in the US retail market for crypto trading. While volatile, trading activity has continued to grow, and while regulatory risks remain, there have been incrementally positive developments in the regulatory landscape.

Upside case **USD 230.00**

Clear regulations for crypto assets emerge with immaterial impact on COIN's operating model reducing investor uncertainty and driving higher valuations. Crypto asset prices rebound, providing a boost to trading revenues.

Downside case **USD 103.00**

Regulatory overhang could worsen, market trading volumes could fail to recover, and fee rates could decline faster than expected. Recent layoffs make hiring new talent more difficult and ambiguity regarding regulation persists, driving down valuations and estimates.

Upside/Downside scenarios

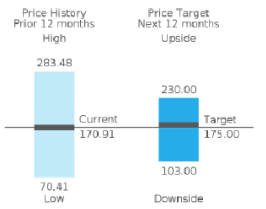
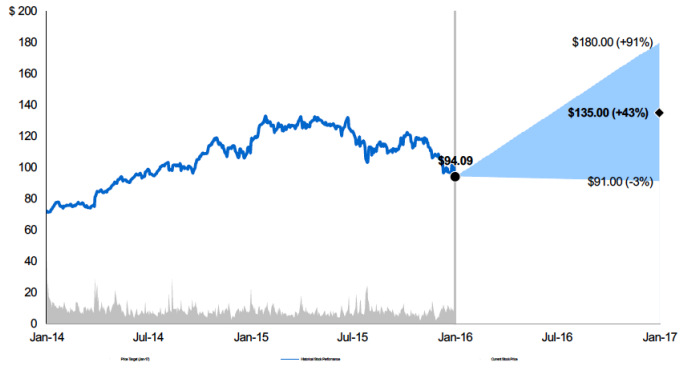


Fig. A.1. **Barclays Report on Coinbase Global Inc.**

Risk Reward

World's Most Valuable Technology Platform Enhanced by New Devices and Services



Source: Thomson Reuters, Morgan Stanley Research

Price Target $135 Derived From Base-Case Scenario.

Bull $180
18x Bull Case FY16e EPS of $10.00

iPhone demand better than feared and investors focus on platform valuation. iPhone units hold up better and ASP grows. Investors focus on Apple's accelerating Services revenue and one billion device installed base. Revenue grows 5% and gross margin remains flat in FY16 driven by Watch and Services. We assume an 18x P/E multiple, in-line with US, large-cap platform companies across industries, or 15x adjusting for Apple's net cash.

Base $135
15x Base Case FY16e EPS of $9.00 or 12x Ex-Cash

Weaker iPhone demand short term but long term opportunities remain. Investors price in weaker iPhone demand this cycle but remain optimistic about iPhone 7 and increasingly new markets like autos. Revenue faces strong foreign currency headwinds and decline 3% in FY16. iPhone units and ASPs fall 7% and 2%. Gross margin contracts 40 bps Y/Y. We assume a 15x P/E multiple, similar to the high-end of where Apple traded in the last year, or 12x adjusting for Apple's $153B of net cash. We believe stable earnings and FCF, despite FY16 being the weaker half of the two year iPhone cycle, will help the stock re-rate closer to a platform multiple, reflective of recurring revenue from Apple's loyal customer base.

Bear $91
11x Bear Case FY16e EPS of $8.25

Both iPhone units and ASPs disappoint. Revenue declines 8% and gross margins contract one point in FY16 mainly due to iPhones, which is slightly offset by growth in Watches and Services. Apple continues to invest in future products and services, driving significant negative operating leverage. P/E multiple falls to 11x or 8x after adjusting for Apple's net cash balance.

Investment Thesis

- Apple has the world's most valuable technology platform with over one billion active devices, and we believe it is best positioned to capture more of its users' time in areas such as health, autos and home, as these platforms expand in the Internet of Things computing era. We believe the company deserves a platform multiple of high-teens, which drives our bull case.

Key Debates

- *Can Apple grow revenue and EPS?* Yes, at a low teens pace over time as it takes share in slower growth smartphone and tablet markets with larger screens and new services. However, new product categories (like Watch), services (Apple Music), and partnerships (HealthKit, HomeKit, CarPlay) could further boost growth with the addition of TV and autos longer-term.
- *Can Apple accelerate innovation?* Yes, FY16 is the fourth year in a row R&D growth should outpaced revenue growth. We see Watch and Apple TV as important barometers of the company's innovation capabilities under the leadership of Tim Cook. We are also encouraged by recent additions to Apple's management team, which expand leadership in key areas like retail, design, health, digital content, and recently autos.

Potential Catalysts

- Better than expected iPhone demand, potentially boosted by a new 4", lower priced version and the iPhone Upgrade Program and other installment and leasing programs, which should shorten refresh cycles
- New products (Watch, MacBook, iPad Pro, Apple TV) and services (Music, Pay, potentially TV subscriptions) drive growth and increase in mix
- Expanding the platform to new industries, for example in healthcare, autos or homes, drives "halo effect" across Apple's businesses
- Expanding points-of-sale, especially in emerging markets like China, India and Brazil

Risks to Achieving Price Target

- Weak global consumer spending and strong US Dollar create headwinds
- Maturing markets, and Android competition in smartphones and tablets
- Lack of traction with new product categories

Fig. A.2. **Morgan Stanley Report on Apple Inc.**

EQUITY RESEARCH

Capital Markets

RBC Capital Markets, LLC
Ben Hendrix (Analyst)
(615) 372-1323, ben.hendrix@rbccm.com
Michael Murray, CFA (Senior Associate)
(615) 372-1321, michael.murray@rbccm.com

October 7, 2024

Humana Inc.

Modeling a 15% EPS decline in 2026 on star bonus headwinds, but maintaining OP

Our view: Following Humana's 8-K last week confirming a star ratings decline based on preliminary MA performance data, we are introducing our 2026 adjusted EPS estimate of $16.99, which assumes an approximate $8.00 EPS impact from lost bonus revenue in 2026. We are lowering our price target to $265 (from $400), reflecting a PE multiple of 15.6x our 2026 estimate, which we believe gives HUM credit for mitigation initiatives for 2027 and leaves some upside for potential success with pending appeals. With 12% implied upside to our PT, we are maintaining OP.

Key points:

- **Following Humana's confirmation last week of its expected Medicare Advantage Star Ratings miss for the 2025 plan year (2026 bonus payment year), we are introducing our 2026 adjusted EPS estimate of $16.99, which reflects our initial estimate for an approximate $8.00 adjusted EPS headwind related to lost bonus payment revenue in 2026.** We are lowering our price target to $265 (from $400), which represents 12% implied upside to Monday's close. We believe our price target reflects the topline deficit in 2026 related to the stars bonus headwind, with modest upside from current levels related to management's initiatives to improve performance on star metrics ahead of the 2027 bonus year (as discussed in our reference notes from Humana's call with sell-side analysts published on October 2, 2024, available here), as well as what we believe is a modest chance the company is successful on appeal of the star rating miss.
- **In this note we lay out assumptions behind our 2026 adjusted EPS estimate of $16.99 and provide sensitivity analysis around key factors management noted would ultimately determine the net impact of the lower ratings.** These include the Medicare Advantage funding environment and traction mitigation efforts, which we believe could include reduction in G&A costs.
- **While we had not published a 2026 adjusted EPS estimate for HUM prior to its star ratings 8-K last week, under our base case assumptions and in the absence of a stars miss, we believe an appropriate EPS value for 2026 would have been in the ~$25.00 range. Base case operating assumptions include modest MA member growth of just over 3% in 2026, following management's expected 10% decline for 2025 on plan design changes.** Our base case scenario for MA funding growth in 2026 is approximately 3%. Our model also contemplated 2026 insurance MCR of roughly 89.1% and an operating cost ratio of ~12%. **Our current $16.99 EPS estimate for 2026 essentially maintains the MA rate update assumption and holds the dollar amount of medical and operating costs constant against our estimate of curtailed stars bonus revenue.**
- **Our upside valuation scenario of $379 assumes a successful appeal of CMS' rating results,** which we believe could result in lost bonus payments for only ~300k MA members (versus ~3.6MM in our base case).

Outperform
NYSE: HUM; USD 236.29
Price Target USD 265.00 ↓ 400.00

WHAT'S INSIDE

☐ Rating/Risk Change	☑ Price Target Change
☐ In-Depth Report	☑ Est. Change
☐ Preview	☑ News Analysis

Scenario Analysis*

Downside Scenario	Current Price	Price Target	Upside Scenario
212.00	236.29	265.00	379.00
↓ 9%		↑ 14%	↑ 62%

*Implied Total Returns

Key Statistics

Shares O/S (MM):	121.3	Market Cap (MM):	28,662
Dividend:	3.54	Yield:	1.5%
Debt to Cap:	44%	Enterprise Val. (MM):	39,633
		Avg. Daily Volume:	2,400,260

RBC Estimates

FY Dec	2023A	2024E	2025E	2026E
EPS, Adj Diluted	26.09	16.00	20.00	16.99
Prev.		15.99		
P/AEPS	9.1x	14.8x	11.8x	13.9x
Revenue	106,374.0	116,065.0	114,348.0	118,593.0

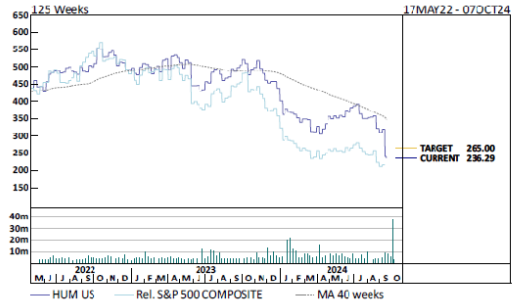
EPS, Adj Diluted	Q1	Q2	Q3	Q4
2023	9.38A	8.94A	7.78A	(0.11)A
2024	7.23A	6.96A	3.29E	(1.49)E
Revenue				
2023	26,742.0A	26,747.0A	26,423.0A	26,462.0A
2024	29,611.0A	29,540.0A	28,374.0E	28,541.0E

All values in USD unless otherwise noted.
Priced as of prior trading day's market close, EST (unless otherwise noted).

Fig. A.3. **RBC Report on Humana Inc.**

Target/Upside/Downside Scenarios

Humana Inc.



Source: Bloomberg and RBC Capital Markets estimates for Target

Valuation

Our price target of $265 reflects a target PE multiple of 15.6x our 2026 adjusted EPS estimate, factoring in some potential for a successful appeal of Humana's recent star rating results. We believe our multiple gives the company some credit for rating remediation efforts, which could support bonus payment recovery in 2027. Management has discussed strategic initiatives to bolster scores across 18 stars measures, 14 of which the company could gain traction within time to affect 2027 bonus payments. While we believe recovery toward the new cut points will be a heavy lift, we cite recovery in CVS' national PPO (contract H5521) to 4.0 stars ahead of the 2024 plan year, from 3.5 into 2023, as an encouraging precedent for quick recovery across a large percentage of MA members. Our 15.6x target multiple represents a one-turn discount to peer UNH, the industry leader by absolute number of Medicare Advantage members. Our price target supports our Outperform rating.

Upside scenario

Our upside valuation scenario of $379 reflects a successful appeal scenario. We believe our 2026 adjusted EPS estimate would see significant improvement in the event of a successful star rating appeal, bringing the three Humana contracts that narrowly miss the 4-star mark back into bonus territory. In this scenario we maintain our base case 15.6x PE multiple assumption for 2026, which continues to represent a one-turn discount to peer UNH.

Downside scenario

Our downside scenario valuation of $212 per share assumes a PE multiple of 12.5x our 2026 adjusted EPS estimate, in line with trading levels as of late September, prior to announcement of the expected stars miss. We believe a pre-announcement multiple on our 2026 estimate would reflect little traction with bonus recovery efforts for 2027.

Investment summary

The company's history of performance in Medicare Advantage and commitment to stable plan design is a key competitive differentiator that solidifies its position as a leader in the Medicare Advantage market, in our view.

Humana continues to build out its primary care and other healthcare services platforms, providing integrated clinical capabilities to improve health outcomes and advance value-based care.

While the elevated utilization headwinds seen in recent quarters are likely to continue in 2024, we are encouraged to see the resetting of EPS targets, which we believe positions the company well for long-term sustainable growth.

We continue to believe Humana's above-industry MA Star Rating profile puts the company on solid competitive footing amid regulatory/reimbursement changes.

Potential catalysts

- Quarterly earnings results.
- Industry M&A activity.
- Inflection points in utilization.

Risks to rating and price target

- Execution risk in predicting and managing medical cost trends.
- Execution risk in network contracting.
- Regulatory risk.
- Reimbursement risk.

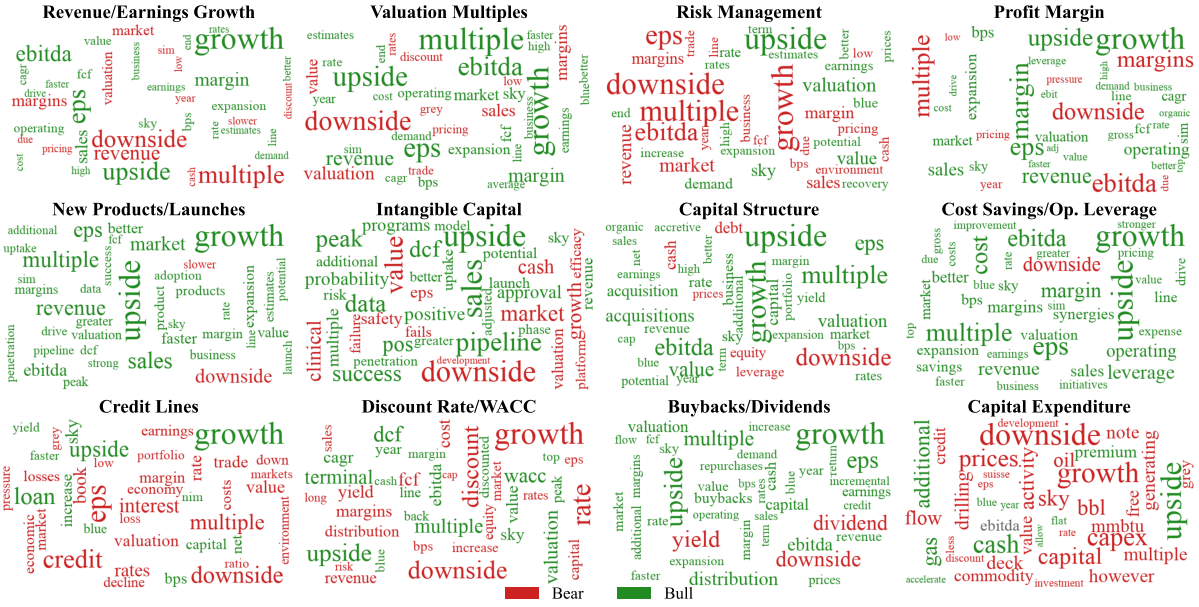


Fig. A.4. **Word Clouds of Firm-Level Sub-Topics: Bear vs. Bull Scenarios**

Each panel displays a word cloud for one of the 12 firm-level sub-topics. Words appearing more frequently in bear scenario narratives are shown in red; words more frequent in bull scenario narratives are shown in green. Word size reflects overall frequency across both scenarios.

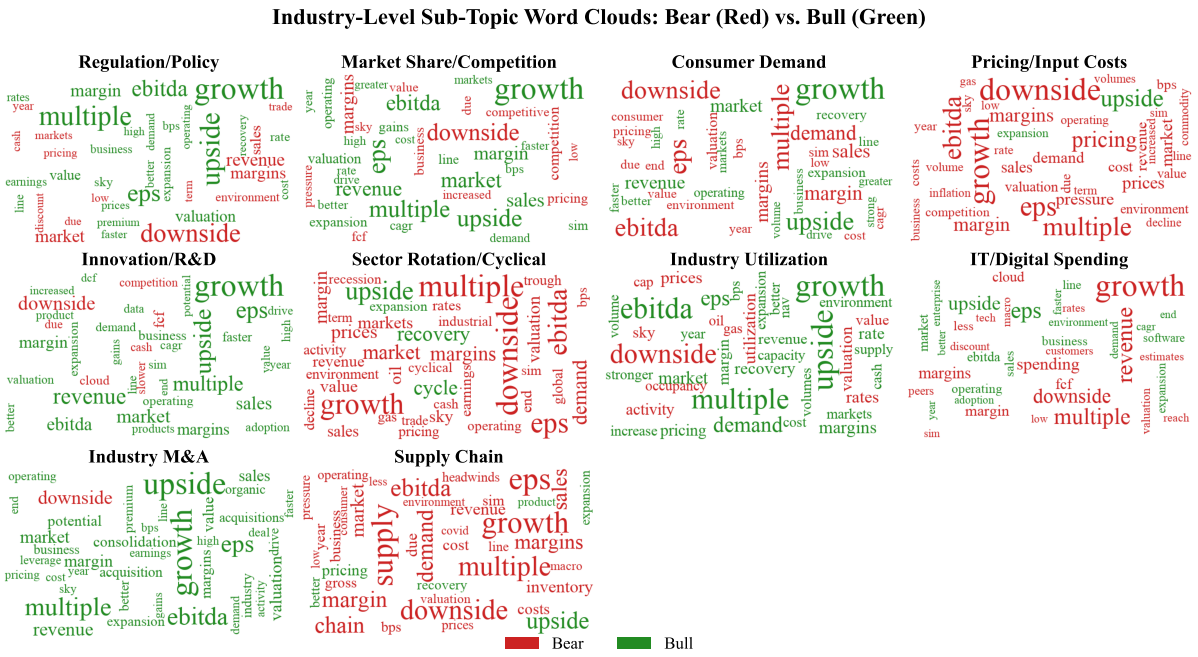


Fig. A.5. **Word Clouds of Industry-Level Sub-Topics: Bear vs. Bull Scenarios**

Each panel displays a word cloud for one of the 10 industry-level sub-topics. Words appearing more frequently in bear scenario narratives are shown in red; words more frequent in bull scenario narratives are shown in green. Word size reflects overall frequency across both scenarios.

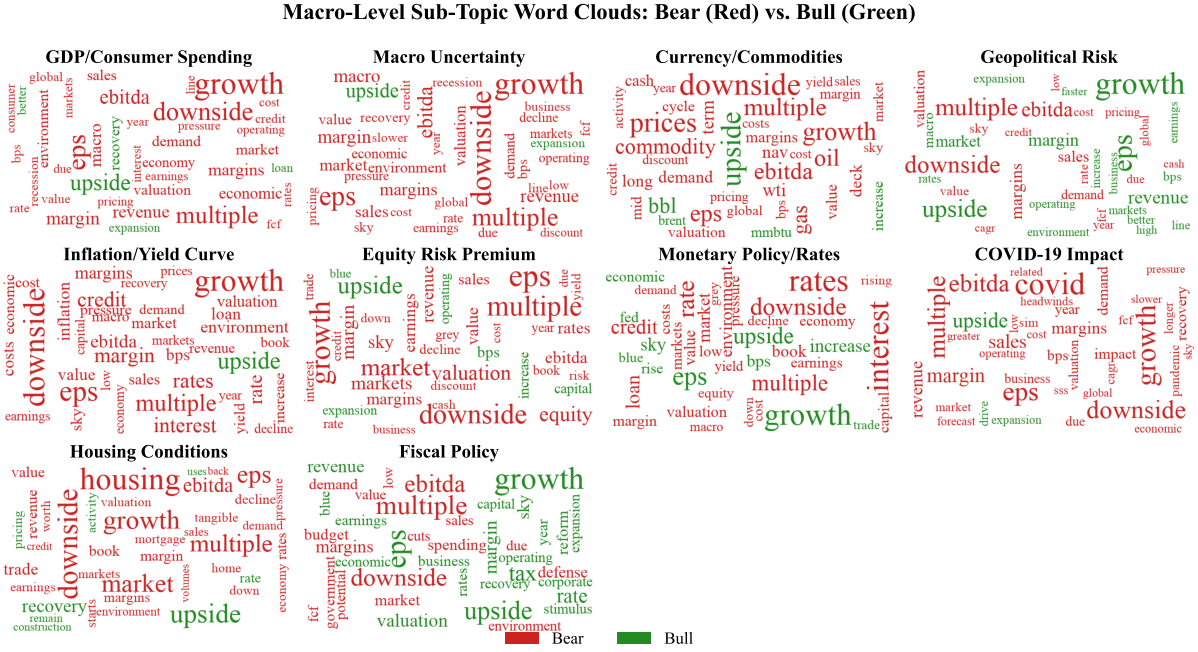


Fig. A.6. **Word Clouds of Macro-Level Sub-Topics: Bear vs. Bull Scenarios**

Each panel displays a word cloud for one of the 10 macro-level sub-topics. Words appearing more frequently in bear scenario narratives are shown in red; words more frequent in bull scenario narratives are shown in green. Word size reflects overall frequency across both scenarios.

Table A.1: **Higher-Moment Expectations and Future Stock Returns**

This table examines whether analyst subjective variance and skewness expectations predict subsequent 12-month cumulative stock returns. The dependent variable is the cumulative log return over the 252 trading days following each analyst report date ($Ret_{t+1,t+12}$). In Panel A, the key independent variable is the analyst variance expectation ($IV^{Analyst}$), controlling for option-implied variance (IV^{Option}) and past realized variance (RV^{Past}). In Panel B, the key independent variable is the analyst skewness expectation ($Skew^{Analyst}$), controlling for option-implied skewness ($Skew^{Option}$) and past realized skewness ($Skew^{Past}$). Column (2) adds the analyst's own return expectation ($r^{Analyst}_{t\to t+12}$) to control for the first moment of the belief distribution. Firm characteristics (market beta, log market equity, log book-to-market, asset growth, gross profitability-to-asset, 12-2 momentum, and 1-month reversal) are included as controls where indicated. Specifications progressively add broker and year-month fixed effects. All independent variables are winsorized at the 1st and 99th percentiles within each year-month. Standard errors are three-way clustered by broker, firm, and year-month; t-statistics are reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Variance Expectations and Future Returns

	$Ret_{t+1,t+12}$				
	(1)	(2)	(3)	(4)	(5)
$IV^{Analyst}_{t+1,t+12}$	-0.225***	-0.242***	-0.132***	-0.077**	-0.081**
	[-4.34]	[-5.57]	[-3.43]	[-2.48]	[-2.61]
$r^{Analyst}_{t\to t+12}$		-0.111**	-0.056	-0.032	-0.029
		[-2.98]	[-1.66]	[-1.02]	[-0.95]
$IV^{Option}_{t+1,t+12}$			-0.062	-0.075**	-0.074**
			[-1.11]	[-2.45]	[-2.44]
$RV^{Past}_{t-12,t}$			0.045	-0.109*	-0.109*
			[0.54]	[-2.06]	[-2.09]
Firm Characteristics	No	No	Yes	Yes	Yes
Broker FE	No	No	No	Yes	No
Year-Month FE	No	No	No	Yes	No
Broker×Year-Month FE	No	No	No	No	Yes
N	180,182	180,182	180,182	180,181	180,119
Adj. R^2	0.019	0.021	0.037	0.173	0.184

Panel B: Skewness Expectations and Future Returns

	$Ret_{t+1,t+12}$				
	(1)	(2)	(3)	(4)	(5)
$Skew^{Analyst}_{t+1,t+12}$	-0.028***	-0.041***	-0.010**	-0.007	-0.007
	[-3.37]	[-4.64]	[-2.26]	[-1.80]	[-1.65]
$r^{Analyst}_{t\to t+12}$		-0.100**	-0.026	-0.003	0.002
		[-2.76]	[-0.67]	[-0.09]	[0.05]
$Skew^{Option}_{t+1,t+12}$			-0.116***	-0.086***	-0.087***
			[-5.40]	[-5.66]	[-5.79]
$Skew^{Past}_{t-12,t}$			-0.012***	-0.010***	-0.010***
			[-3.41]	[-4.18]	[-4.27]
Firm Characteristics	No	No	Yes	Yes	Yes
Broker FE	No	No	No	Yes	No
Year-Month FE	No	No	No	Yes	No
Broker×Year-Month FE	No	No	No	No	Yes
N	180,182	180,182	180,182	180,181	180,119
Adj. R^2	0.002	0.003	0.039	0.170	0.181

Table A.2: **Mixture Model: Predicting Realized Variance and Skewness**

This table replicates Table 2 using the two-component mixture of normals instead of the binomial tree model. The dependent variable in Panel A is annualized realized variance ($RV^{Realized}_{t+1,t+12}$), computed from daily log returns over the 252 trading days following each analyst report date; in Panel B, it is realized skewness ($Skew^{Realized}_{t+1,t+12}$), the standardized third central moment over the same window. $IV^{Mixture}$ and $Skew^{Mixture}$ are the subjective variance and skewness expectations derived from a two-component Gaussian mixture that augments the binomial tree with a continuous density component. IV^{Option} and $Skew^{Option}$ are the model-free option-implied variance and skewness of Bakshi, Kapadia, and Madan (2003), computed from the volatility surface at a 365-day horizon. RV^{Past} and $Skew^{Past}$ are the realized variance and skewness over the prior 252 trading days. Firm characteristics (market beta, log market equity, log book-to-market, asset growth, and gross profitability-to-asset) are included as controls where indicated. Specifications progressively add broker and year-month fixed effects. All independent variables are winsorized at the 1st and 99th percentiles within each year-month. Standard errors are three-way clustered by broker, firm, and year-month; t-statistics are reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is 2007 to 2024.

Panel A: Predicting Future Realized Variance

	$RV^{Realized}_{t+1,t+12}$			
	(1)	(2)	(3)	(4)
$IV^{Mixture}_{t+1,t+12}$	0.428***	0.138***	0.120***	0.122***
	[12.93]	[8.35]	[8.46]	[8.50]
$IV^{Option}_{t+1,t+12}$		0.350***	0.333***	0.331***
		[12.12]	[10.87]	[10.93]
$RV^{Past}_{t-12,t}$		0.158**	0.304***	0.306***
		[2.94]	[6.68]	[6.78]
Firm Characteristics	No	Yes	Yes	Yes
Broker FE	No	No	Yes	No
Year-Month FE	No	No	Yes	No
Broker×Year-Month FE	No	No	No	Yes
N	180,206	180,206	180,205	180,143
Adj. R^2	0.179	0.313	0.376	0.382

Panel B: Predicting Future Realized Skewness

	$Skew^{Realized}_{t+1,t+12}$			
	(1)	(2)	(3)	(4)
$Skew^{Mixture}_{t+1,t+12}$	0.073***	0.056***	0.062***	0.063***
	[8.29]	[4.91]	[4.39]	[4.32]
$Skew^{Option}_{t+1,t+12}$		0.018	0.026	0.027
		[0.43]	[0.61]	[0.64]
$Skew^{Past}_{t-12,t}$		0.003	0.002	0.002
		[0.33]	[0.18]	[0.24]
Firm Characteristics	No	Yes	Yes	Yes
Broker FE	No	No	Yes	No
Year-Month FE	No	No	Yes	No
Broker×Year-Month FE	No	No	No	Yes
N	180,206	180,206	180,205	180,143
Adj. R^2	0.001	0.004	0.013	0.020

Table A.3: **Mixture Model: Risk–Return and Skewness–Return Trade-Offs**

This table replicates Table 8 using the two-component mixture of normals instead of the binomial tree model. The dependent variable is the analyst's log return expectation, $r^{Analyst}_{b,i,t\to t+12} = \ln(\text{Price Target}/P_t)$. In Panel A, the key regressor is the mixture variance ($IV^{Mixture}_{t+1,t+12}$); in Panel B, it is the mixture skewness ($Skew^{Mixture}_{t+1,t+12}$). $IV^{Mixture}$ and $Skew^{Mixture}$ are derived from a two-component Gaussian mixture that augments the binomial tree with a continuous density component. Columns (1)–(2) use all analyst reports with broker and year-month fixed effects; column (1) includes only the mixture expectation, and column (2) adds firm characteristics: market beta, log market capitalization (log(ME)), log book-to-market (log(B/M)), asset growth (AG), and gross profitability-to-asset (GPA). Columns (3)–(5) report results estimated separately for the three largest brokers with sufficiently large firm coverage over a long sample period: Barclays, Morgan Stanley, and RBC, each with year-month fixed effects and firm characteristics. Standard errors are three-way clustered by broker, firm, and year-month for columns (1)–(2), and two-way clustered by firm and year-month for columns (3)–(5); t-statistics appear in brackets. All independent variables are winsorized at the 1st and 99th percentiles. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel A: Risk-Return Trade-Off

	$r^{Analyst}_{i,t\to t+12}$				
	(1) All	(2) All	(3) Barclays	(4) Morgan Stanley	(5) RBC
$IV^{Mixture}_{t+1,t+12}$	0.194***	0.151***	0.110	0.048**	0.282***
	[4.30]	[3.72]	[1.64]	[2.56]	[19.43]
β		0.027***	0.029***	0.033***	0.041***
		[3.90]	[3.44]	[3.42]	[5.70]
log(ME)		-0.005	-0.003	0.007**	-0.007***
		[-1.48]	[-0.60]	[2.18]	[-3.80]
log(B/M)		-0.005	0.002	0.006	-0.000
		[-1.38]	[0.54]	[1.36]	[-0.15]
GPA		-0.120***	-0.126***	-0.090***	-0.044***
		[-8.38]	[-6.63]	[-3.88]	[-2.81]
AG		0.032***	0.032***	0.035***	0.035***
		[10.71]	[5.02]	[4.89]	[6.51]
Broker FE	Yes	Yes	No	No	No
Year-Month FE	Yes	Yes	Yes	Yes	Yes
N	180,857	180,857	43,639	32,237	31,395
Adj. R^2	0.183	0.209	0.130	0.108	0.353

Panel B: Skewness-Return Trade-Off

	$r^{Analyst}_{i,t\to t+12}$				
	(1) All	(2) All	(3) Barclays	(4) Morgan Stanley	(5) RBC
$Skew^{Mixture}_{t+1,t+12}$	-0.074***	-0.080***	-0.077***	-0.120***	-0.059***
	[-11.47]	[-12.46]	[-19.98]	[-22.83]	[-16.79]
β		0.049***	0.045***	0.048***	0.058***
		[7.43]	[6.61]	[5.62]	[7.38]
$\log(ME)$		-0.021***	-0.013***	-0.006*	-0.028***
		[-3.92]	[-5.00]	[-1.89]	[-9.44]
$\log(B/M)$		-0.006	0.003	0.005	-0.008**
		[-1.56]	[0.74]	[1.53]	[-1.99]
GPA		-0.152***	-0.111***	-0.089***	-0.135***
		[-6.35]	[-5.55]	[-4.57]	[-6.40]
AG		0.035***	0.031***	0.037***	0.038***
		[10.23]	[5.22]	[5.41]	[6.37]
Broker FE	Yes	Yes	No	No	No
Year-Month FE	Yes	Yes	Yes	Yes	Yes
N	180,857	180,857	43,639	32,237	31,395
Adj. R^2	0.159	0.256	0.245	0.244	0.271

Table A.4: **Three-State Model: Predicting Realized Variance and Skewness**

This table replicates Table 2 using the three-state model instead of the binomial tree. The three-state model assigns separate probabilities p_{up}, p_{down}, and $p_{base} = 1 - p_{up} - p_{down}$ to the bull, bear, and base scenarios, respectively, without constraining $p_{up} + p_{down} = 1$. Probabilities are estimated at the broker level from an expanding window of past reports, using two moment conditions: (i) the probability-weighted expected price equals the base-case price, and (ii) the probability-weighted variance equals the average realized variance of past reports from the same broker. To avoid look-ahead bias, estimation uses reports from years $\leq t-2$ and applies the estimated probabilities to year-t reports. The dependent variable in Panel A is annualized realized variance ($RV^{Realized}_{t+1,t+12}$), computed from daily log returns over the 252 trading days following each report date; in Panel B, it is realized skewness ($Skew^{Realized}_{t+1,t+12}$), the standardized third central moment over the same window. IV^{3State} and $Skew^{3State}$ are the subjective variance and skewness expectations derived from the three-state model. IV^{Option} and $Skew^{Option}$ are the model-free option-implied variance and skewness of Bakshi, Kapadia, and Madan (2003), computed from the volatility surface at a 365-day horizon. RV^{Past} and $Skew^{Past}$ are the realized variance and skewness over the prior 252 trading days. Firm characteristics (market beta, log market equity, log book-to-market, asset growth, and gross profitability-to-asset) are included as controls where indicated. Specifications progressively add broker and year-month fixed effects. All independent variables are winsorized at the 1st and 99th percentiles within each year-month. Standard errors are three-way clustered by broker, firm, and year-month; t-statistics are reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Predicting Future Realized Variance

	$RV^{Realized}_{t+1,t+12}$			
	(1)	(2)	(3)	(4)
$IV^{3State}_{t+1,t+12}$	0.566***	0.199***	0.173***	0.175***
	[17.98]	[9.49]	[9.48]	[9.68]
$IV^{Option}_{t+1,t+12}$		0.334***	0.320***	0.318***
		[11.19]	[10.14]	[10.19]
$RV^{Past}_{t-12,t}$		0.146**	0.292***	0.294***
		[2.73]	[6.39]	[6.49]
Firm Characteristics	No	Yes	Yes	Yes
Broker FE	No	No	Yes	No
Year-Month FE	No	No	Yes	No
Broker×Year-Month FE	No	No	No	Yes
N	169,352	169,352	169,352	169,320
Adj. R^2	0.201	0.320	0.379	0.385

Panel B: Predicting Future Realized Skewness

	$Skew^{Realized}_{t+1,t+12}$			
	(1)	(2)	(3)	(4)
$Skew^{3State}_{t+1,t+12}$	0.077***	0.067***	0.071***	0.071***
	[4.85]	[3.96]	[3.96]	[3.79]
$Skew^{Option}_{t+1,t+12}$		0.035	0.032	0.033
		[0.81]	[0.74]	[0.77]
$Skew^{Past}_{t-12,t}$		0.003	0.002	0.002
		[0.28]	[0.16]	[0.20]
Firm Characteristics	No	Yes	Yes	Yes
Broker FE	No	No	Yes	No
Year-Month FE	No	No	Yes	No
Broker×Year-Month FE	No	No	No	Yes
N	169,352	169,352	169,352	169,320
Adj. R^2	0.001	0.004	0.013	0.019

Table A.5: **Three-State Model: Risk–Return and Skewness–Return Trade-Offs**

This table replicates Table 8 using the three-state model instead of the binomial tree. The three-state model assigns separate probabilities p_{up}, p_{down}, and $p_{base} = 1 - p_{up} - p_{down}$ to the bull, bear, and base scenarios, respectively. Probabilities are estimated at the broker level from an expanding window of past reports, using two moment conditions: (i) the probability-weighted expected price equals the base-case price, and (ii) the probability-weighted variance equals the average realized variance of past reports from the same broker. To avoid look-ahead bias, estimation uses reports from years $\leq t-2$ and applies the estimated probabilities to year-t reports. The dependent variable is the analyst's log return expectation, $r^{Analyst}_{b,i,t\to t+12} = \ln(\text{Price Target}/P_t)$. In Panel A, the key regressor is the three-state variance ($IV^{3State}_{t+1,t+12}$); in Panel B, it is the three-state skewness ($Skew^{3State}_{t+1,t+12}$). Columns (1)–(2) use all analyst reports with broker and year-month fixed effects; column (1) includes only the three-state expectation, and column (2) adds firm characteristics: market beta, log market capitalization (log(ME)), log book-to-market (log(B/M)), asset growth (AG), and gross profitability-to-asset (GPA). Columns (3)–(5) report results estimated separately for the three largest brokers with sufficiently large firm coverage over a long sample period: Barclays, Morgan Stanley, and RBC, each with year-month fixed effects and firm characteristics. Standard errors are three-way clustered by broker, firm, and year-month for columns (1)–(2), and two-way clustered by firm and year-month for columns (3)–(5); t-statistics appear in brackets. All independent variables are winsorized at the 1st and 99th percentiles. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel A: Risk-Return Trade-Off

	$r^{Analyst}_{i,t\to t+12}$				
	(1) All	(2) All	(3) Barclays	(4) Morgan Stanley	(5) RBC
$IV^{3State}_{t+1,t+12}$	0.234***	0.177***	0.096	0.028	0.330***
	[4.54]	[3.80]	[1.26]	[1.21]	[15.55]
β		0.026***	0.024***	0.039***	0.038***
		[3.71]	[2.84]	[4.20]	[5.14]
$\log(ME)$		-0.006	-0.005	0.006*	-0.009***
		[-1.73]	[-1.18]	[1.78]	[-3.88]
$\log(B/M)$		-0.005	0.001	0.005	-0.002
		[-1.60]	[0.18]	[1.26]	[-0.57]
GPA		-0.124***	-0.140***	-0.098***	-0.052***
		[-8.35]	[-6.58]	[-4.28]	[-3.11]
AG		0.033***	0.030***	0.037***	0.036***
		[11.39]	[4.54]	[5.02]	[6.74]
Broker FE	Yes	Yes	No	No	No
Year-Month FE	Yes	Yes	Yes	Yes	Yes
N	169,971	169,971	37,894	32,012	31,147
Adj. R^2	0.179	0.207	0.118	0.098	0.338

Panel B: Skewness-Return Trade-Off

	$r^{Analyst}_{i,t\to t+12}$				
	(1) All	(2) All	(3) Barclays	(4) Morgan Stanley	(5) RBC
$Skew^{3State}_{t+1,t+12}$	-0.113***	-0.110***	-0.103***	-0.120***	-0.095***
	[-27.15]	[-23.56]	[-22.78]	[-26.49]	[-22.88]
β		0.037***	0.027***	0.032***	0.050***
		[5.18]	[3.97]	[3.85]	[6.55]
$\log(ME)$		-0.020***	-0.014***	-0.002	-0.026***
		[-3.72]	[-5.42]	[-0.84]	[-9.41]
$\log(B/M)$		-0.005	0.002	0.007*	-0.005
		[-1.33]	[0.64]	[1.87]	[-1.48]
GPA		-0.141***	-0.102***	-0.076***	-0.123***
		[-6.03]	[-4.81]	[-3.82]	[-6.22]
AG		0.030***	0.026***	0.032***	0.034***
		[9.36]	[4.43]	[4.75]	[5.98]
Broker FE	Yes	Yes	No	No	No
Year-Month FE	Yes	Yes	Yes	Yes	Yes
N	169,971	169,971	37,894	32,012	31,147
Adj. R^2	0.223	0.299	0.273	0.261	0.318

A.2. LLM-Based Topic Extraction and Classification

This appendix describes the large language model (LLM) procedure used to map analyst scenario discussions into topic probability distributions, as described in Section 4. Our approach relates to the growing literature that uses LLMs to extract structured information from financial text (Bastianello, Decaire, and Guenzel, 2024, 2025; Ke, 2024). The procedure consists of four steps: open-ended topic extraction, taxonomy construction, human validation, and structured topic classification.

In the first step, we use Gemini 2.0 Flash to extract free-form topics from each scenario text. For each of the three scenarios (bull, bear, and base) in every report, the model identifies all economically meaningful topics discussed in the text. Each topic is assigned to an entity category (firm, industry, or macroeconomy) and a valuation channel (e.g., earnings/cash flow, discount rate, or valuation), and the model also returns a short supporting text snippet. Because this extraction is open-ended and does not rely on a pre-specified taxonomy, it allows the model to recover the full range of themes present in analyst narratives. The complete prompt is reproduced below.

In the second step, we consolidate the raw extracted topics into a unified taxonomy. The open-ended extraction yields approximately 493,000 topic mentions, including more than 150,000 unique topic phrases at the firm level, 32,000 at the industry level, and 11,000 at the macroeconomy level. Because topics differ substantially across these three entity categories, we construct the taxonomy separately within each entity level. Specifically, we use semantic text embeddings to represent topic phrases and apply unsupervised clustering to group closely related expressions. We select the number of clusters within each entity category using the silhouette score, while also considering the economic interpretability of the resulting clusters. We then assign economically meaningful labels to the resulting groups by inspecting the most representative topic phrases within each cluster. This procedure yields 25 firm-level clusters, 15 industry-level clusters, and 6 macroeconomy-level clusters.

For each of the three scenarios (BASE, BULLISH, BEARISH), analyze the scenario discussion from an equity analyst report. Your task is to extract all topics in this scenario.

1. **Topic Identification.** Each identified topic must be specific, economically meaningful, and tied to a single factor influencing valuation.

2. **Entity Identification (Choose Exactly One).** Assign each topic to one of the following entities:

- firm — company-specific factors
- industry — industry or sector conditions
- macroeconomy — macroeconomic conditions

If multiple entities are plausible, select the one best supported by the context.

3. **Channel Identification (Choose Exactly One).** Assign each topic to the correct valuation channel based on context:

- earnings/cash flow: sales — related to sales or revenues
- earnings/cash flow: costs — related to costs or operating expenses
- earnings/cash flow: margins — related to profitability margins
- earnings/cash flow: earnings — related to earnings, EBITDA, EBIT, or cash flows
- discount rate — related to discount rate, WACC, equity beta, idiosyncratic risk, credit rating, systematic risk, equity risk premium, market risk premium, interest rate, risk-free rate, or treasury yield
- valuation — related to valuation terms such as "undervalued," "overvalued," or "fair-valued"
- other — if none of the above apply

4. **Evidence Snippet.** Identify a 5-word snippet from the text to capture where the topic is first discussed.

5. **Output Format (Strict JSON).** For each scenario, return a JSON list. Each topic:

```
{"topic": "...", "entity": "...", "channel": "...", "evidence_snippet": "..."}
```

If a scenario has no meaningful topics or its text is empty, return an empty list [] for that scenario.

In the third step, we manually review and validate the resulting taxonomy. Following this review, the initial clusters are refined and consolidated into a final taxonomy containing 32 sub-topics. We verify that each sub-topic is economically distinct and interpretable, that entity

assignments are internally consistent, and that the taxonomy provides broad coverage of the topics appearing in the raw extractions.

In the fourth step, we use Gemini 2.0 Flash to classify each scenario text into the finalized 32-topic taxonomy. Each scenario text is processed independently. The model receives a system instruction defining the taxonomy and returns a structured JSON object containing probability distributions over the three entity categories and their associated sub-topics. These probabilities provide our final scenario-level topic measures. An abbreviated version of the classification prompt is reproduced below.

> You are a financial analyst assistant specializing in equity research reports. Your task is to analyze short text excerpts from analyst scenario descriptions (bull / base / bear case price justifications) and assign probability distributions that reflect the main topics discussed.
>
> **Entity Categories.** Classify the text across three mutually exclusive entity levels: *firm* (company operations, financials, strategy, valuation), *industry* (sector dynamics, competitive landscape), *macro* (interest rates, GDP, inflation, geopolitics, etc.).
>
> **Sub-Topics.** *Firm (12):* Revenue/Earnings/FCF Growth, Discount Rate/Cost of Capital, Multiple/Peer Group Valuation, Profit Margin, Cost Savings/Operating Leverage, Capital Expenditure, Intangible Capital, New Launches/Product Dynamics, Capital Structure, Buybacks/Dividend, Credit Lines, Risk Assessment and Management. *Industry (10):* Sector Rotation/Cyclical Dynamics, Market Share/Competition, Industry Consolidation (M&A), Industry Utilization, Supply Chain, Pricing Pressure/Input Cost, Consumer Demand, Innovation and R&D, IT & Digital Spending, Regulation/Policy. *Macro (10):* GDP Growth/Unemployment/Consumer Spending, Fiscal Policy/Government Spending, Inflation/Yield Curve/Credit Condition, Monetary Policy/Interest Rates, Equity Market/Risk Premium, Housing Conditions, Currency/Commodity Prices, COVID-19 Pandemic Impact, Macro Uncertainty, Geopolitical Risk. [Each sub-topic includes a detailed description and examples of relevant language patterns.]
>
> **Output Rules.** Return only valid JSON with probabilities over the three entity categories and their associated sub-topics.

Each sub-topic in the full prompt is accompanied by a short description and illustrative language patterns. The complete prompt is available from the authors upon request.