Mental Models and Financial Forecasts*

Francesca Bastianello   Paul H. Décaire   Marius Guenzel†

April 18, 2025

Abstract

We uncover financial professionals' mental models—the reasoning they use to explain their quantitative forecasts. We organize our analysis around a framework of top-down attention, where analysts endogenously choose both a valuation method and how to allocate attention across variables. Using the near-universe of 1.6 million equity analyst reports, we collect the valuation methods analysts adopt to compute their price targets. To measure attention, we then prompt large language models (LLMs) on a subset of 110,000 reports to extract 4.8 million lines of reasoning—each combining a topic, valuation channel, time horizon, and sentiment. To validate the reliability of our output, we introduce a multi-step LLM prompting strategy and new diagnostic tools. We document four main findings. (1) Analysts exhibit sparse mental representations, focusing on a limited set of topics, that are primarily related to top-line items, and forward-looking. (2) The choice of valuation methods and topic focus is closely linked. (3) There is substantial disagreement among analysts, and differences in attention weights to firm-specific variables are a bigger source of disagreement than differences in valuation weights on those variables. (4) Lastly, variation in mental models aligns with key asset pricing patterns both in the time-series and in the cross-section.

Keywords: Mental Models, Professional Forecasters, Large Language Models, Beliefs

*First Version: October 30, 2024. This Version: April 18, 2025. All errors and omissions are our own.

†Bastianello is at the University of Chicago, Booth School of Business. Décaire is at the W. P. Carey School of Business at Arizona State University. Guenzel is at the Wharton School of the University of Pennsylvania.
We thank Sam Hanson, Alex Imas, Chad Kendall, Andrei Shleifer, and seminar and conference participants at the W.P. Carey School of Business at Arizona State University, Harvard Business School, McGill University Desautels Faculty of Management, the 2024 ASU-UA Junior Conference, the 1st Workshop on LLMs and Generative AI for Finance, the 14th ITAM Finance Conference, the 2025 Adam Smith Workshop, and the MIT Sloan Junior Finance Conference for useful comments. Bastianello gratefully acknowledges support from the Liew Fama Junior Faculty Fellowship, the Fama Faculty Fellowship, and the Fama-Miller Center at the University of Chicago Booth School of Business. Décaire gratefully acknowledges financial support from the Center for Responsible Investing at Arizona State University. Guenzel gratefully acknowledges financial support from the Jacobs Levy Equity Management Center for Quantitative Financial Research. Contact information: francesca.bastianello@chicagobooth.edu, paul.decaire@asu.edu, and mguenzel@wharton.upenn.edu.

1 Introduction

How do people form expectations in financial markets? According to the rational expectations benchmark, individuals use all available information to produce statistically optimal forecasts. Early research challenged this view with largely indirect evidence and by documenting market anomalies that are difficult to reconcile with fully rational expectations.[1] Over the last few decades, progress has accelerated with the growing availability of survey data, which allows researchers to directly observe individuals' quantitative forecasts. However, how these expectations are formed remains an open question, partly due to limited measurement. To better understand how market participants mentally represent firms, and how these representations might ultimately drive investor behavior, this paper uses sell-side analysts' written reports to measure the *reasoning* behind their forecasts.

Specifically, we study how analysts reason when valuing stocks. Valuing a firm is an inherently hard problem for several reasons. First, analysts operate in a high-dimensional environment. There are many pieces of information they may wish to attend to, but given their limited attention, they must engage in some form of dimensionality-reduction and choose which variables to attend to, and which ones to neglect. Second, firm valuation is also computationally complex. Applying the present value relationship exactly—that prices equal the present value of all expected future dividends—would require forecasting an infinite stream of future cash-flows and discount rates. In practice, however, this is not how analysts apply this formula. Instead, they rely on simplifying assumptions that give rise to different valuation methods.

To formalize these processes, we define a mental model to be a combination of attention weights and valuation methods, and study how analysts tackle the dual challenge of dimensionality reduction and computational tractability by exploring three sets of questions. First, how do analysts allocate attention and choose valuation methods? Second, how do these choices shape their quantitative forecasts? Third, how do these mental models map onto

[1] See De Bondt and Thaler (1990), Vissing-Jorgensen (2003), Bacchetta et al. (2009), Malmendier and Nagel (2011), Case et al. (2012), Amromin and Sharpe (2014), Greenwood and Shleifer (2014), Bordalo et al. (2020b), Bordalo et al. (2020a), Giglio et al. (2021), Nagel and Xu (2022) among others, and Adam and Nagel (2023) for a review.

observed asset pricing anomalies?

To answer these questions we begin by introducing a conceptual framework of top-down attention, where analysts endogenously choose both attention allocation and valuation methods based on the relevance of variables for valuation and the cost of acquiring information about them. This framework suggests that analysts should pay more attention to variables that are more important, more volatile, and cheaper to learn about. Moreover, the way analysts simplify the information environment through their choice of valuation method should be tightly linked to the information they attend to. Finally, disagreement across analysts may arise either from differing valuation methods or from differences in attention.

The main challenge in testing these theories is that we generally do not observe what investors attend to and how they simplify their valuation task. To overcome these challenges, we leverage the fact that sell-side equity analysts not only publish quantitative forecasts—which have been widely studied in empirical asset pricing—but also produce accompanying reports in which they directly explain the key arguments underlying those forecasts.

We use the near-universe of 1.6 million equity reports from analysts at 43 brokerage houses over a 24-year period to extract forecasters' valuation methods and associated price targets, and we then prompt large language models (LLMs) to extract more than 4.8 million separate lines of reasoning from the text portions of a subset of 110,000 reports. Each line of reasoning is characterized by a topic the analyst may be discussing, together with its associated valuation channel, time outlook, and sentiment. Our resulting dataset links analysts' mental models to their quantitative forecasts for a sample that closely aligns with commonly used commercial databases—such as the Institutional Brokers' Estimate System (IBES)—in firm coverage and characteristics.

We uncover four sets of facts that are consistent with our model of top-down attention. First, analysts exhibit sparse mental representations, and focus on a limited set of arguments, which are predominantly forward-looking and related to top-line items. Second, analysts' attention allocation is tightly linked to their choice of valuation method. Even when covering the same firm at the same point in time, analysts using different valuation methods attend to different pieces of information, leading to distinct representations of the same underlying firm. Third, while both variation in valuation methods and variation in attention drive disagreement, differences in attention to firm-related topics account for the biggest source of disagreement. Finally, we end the paper with a discussion on forecast errors and asset pricing implications, and show how various features of analysts' mental models correlate with key asset pricing patterns both in time-series and in the cross-section.

Conceptual framework. To lay the foundations for our analysis, we start by introducing a conceptual framework built around a simple and standard model of top-down attention. The true value of a stock is determined by a large number of variables, which differ in the weight that they receive for valuation. These variables may include accounting metrics, such as sales and costs at different horizons, as well as intangible factors, such as brand value or customer loyalty. The model has two key components. First, analysts do not directly observe any of the variables but instead receive noisy signals about them. The more attention analysts pay to a given variable, the more precise the signal they obtain. Second, analysts have access to a set of valuation methods, each placing different weights on these variables. For example, we may think of discounted cash flow (DCF) models as assigning greater weight to long-run outcomes, while more reduced-form multiples approaches may emphasize variables associated with relevant comparable. Analysts then choose valuation methods and attention weights to minimize mean squared forecast error subject to a budget constraint on attention.

Given this framework, we first show that if analysts were to use the true valuation function, they would allocate more attention to variables that are more relevant, more volatile, and less costly to acquire information about. Second, we show that when analysts use a valuation method that differs from the true valuation function, they tilt their attention towards the variables that are weighed more in their valuation method. Third, we show that analysts will not necessarily choose the true valuation function, as they face a bias-variance trade-off. The optimal valuation method minimizes discrepancies along variables that are important, while it may allow for discrepancies along less important ones, as this can enable a more efficient allocation of attention. For example, if the true valuation method requires attending to a particularly costly variable, the analyst may be better off neglecting that variable altogether in order to reallocate attention to less costly variables. Finally, given this framework, disagreement across analysts can arise from differences in either valuation methods or attention allocation.

Data collection and methodological contribution. To test the predictions of our model, we construct a dataset that allows us to directly measure both what information analysts attend to, and what simplifying rules they use for valuation. To impose structure on analysts' attention, we draw on the key dimensions of valuation that have been central to debates in empirical asset pricing. For each topic an analyst discusses, we collect information on its associated valuation channel (cash-flows or discount-rates), time outlook (past, present, near-future, or distant future), and sentiment (positive, negative, or neutral). This provides us with a nuanced and multi-dimensional view of what analysts focus on.

To collect this information from equity analyst reports we use large language models (LLMs). One of the advantages of this approach is that the resulting output is interpretable.[2] However, extracting information from large bodies of text also presents two main challenges: ensuring the completeness and richness of the extracted content, and maintaining the stability and reproducibility of the outputs. In this respect, our methodological contribution is twofold. First, we show that when applied to lengthier, multi-page documents—such as equity reports—naïve prompting strategies that ask LLMs to extract information in a single step often yield incomplete outputs. This limitation arises in part from LLMs' tendency to focus disproportionately on the beginning of the input. To address this, we introduce a refined LLM approach in which we *iteratively* provide sub-segments of each report. After processing all segments using this "chunking" method, we then present the LLM with both the full document and the extracted content in a final step to validate the selected topics. This multi-step prompting strategy substantially outperforms the single-step approach in both the number and comprehensiveness of individual lines of reasoning extracted.[3]

Relatedly, we introduce diagnostic tools to show that the output of naïve single-step prompts is unstable, while the output of our multi-step prompt is highly stable. We repeatedly feed a subset of reports to the LLM using both prompting strategies and calculate argument

[2] We discuss the limitations of applying more traditional approaches in our context in Section 3.2. For example, while Latent Dirichlet Allocation (LDA) scales efficiently to thousands or even millions of documents, the resulting output typically requires additional post-hoc inference to assign economic meaning. In contrast, human classifications are more interpretable but lack the scalability required for analyzing text at this magnitude.

[3] Moreover, we show that the additional arguments captured by our multi-step prompt (and omitted by naïve prompts) are strongly predictive of forecasters' quantitative beliefs. This suggests that the additional information we gather via our multi-step strategy is in fact important.

consistency—that is, the likelihood that a given argument is extracted across multiple runs. Unlike naïve one-step approaches, the output of our multi-step approach is highly stable.

Sparse mental representations. Having constructed our dataset, and guided by our theoretical framework, we turn to the key results of our paper. The first fact that we establish is that analysts' mental representations are sparse. In each report, analysts attend to only 18 out of 139 topics, illustrating how their attention allocation in indeed selective.[4] On average, fundamental drivers of cash flow generation, such as customer demand, market share, and cost structure, receive persistent attention over time, while topics like public health, taxation, and inflation are transient and receive attention when they significantly deviate from their steady state levels. There is also variation across industries, once again reflecting the relative importance of different variables in the cross-section. For example, customer demand and retention are especially important in the information sector, while pricing power and the regulatory environment have a much more prominent role for utilities. Finally, at a more aggregate level, the picture is even starker. Analysts predominantly focus on firm-specific topics (75–80%) and top-line items (40–50% of arguments), and their arguments are mostly forward-looking (55-60%). Moreover, sentiment and attention to top-line items is highly pro-cyclical, while attention to discount rates and macro-related topics is counter-cyclical.

Valuation methods and attention are tightly interlinked. The second prediction of our model is that the choice of valuation methods and attention allocation are tightly interlinked. Consistent with this prediction, we find that using the same valuation method is associated with significantly more similar topic focus, as well as greater joint congruence in topic selection, valuation channel, time outlook, and sentiment.

Moreover, while on average 35% of reports use discounted cash flow (DCF) methods, a nearly equal share use price- to-earnings (P/E) multiples, and about 65% use at least one multiples-based approach, there is substantial variation.[5] For example, we find that DCF methods—structured to emphasize future cash flows—are, relatively speaking, more commonly used for small, young, and growth firms, with analysts' reasoning centered on

[4] We present a series of tests to show that topics omitted from textual discussions are outside forecasters' mental models, rather than excluded due to space constraints in reports.

[5] Our extraction of valuation method information allows for more than one valuation methods to be recorded within a single report, which explains why the shares describe above sum to more than 100%.

long-run topics such as discount rates, innovation, and corporate investment. In contrast, multiples-based approaches are more frequently applied to large, mature, and value firms, where analysts tend to focus on product- and customer-related topics.[6]

From mental models to quantitative forecasts and disagreement. Next, we explore how differences in reasoning translate into quantitative forecasts. Both differences in attention and differences in valuation methods contribute to disagreement, but differences in attention play a bigger role. When considering two analysts covering the same firm at the same point in time, they share only 32% of the combined topics used, while they have some overlap in valuation methods used around 70% of the time.

Consistent with these patterns, we find that differences in attention weights contribute more to total disagreement than do differences in valuation weights. Moreover, attention differences to more firm-specific categories—such as payout, product market competition, product, and brand—appear to matter the most, whereas attention differences to more macro-oriented topics—such as interest rates, housing, or government—appear to matter less in explaining disagreement.

Overall, this analysis allows us to make the drivers of disagreement interpretable, and it also suggests that analysts converge more on how they handle computational complexity (what algorithm/valuation method to use), than in how they resolve representational complexity, such as deciding which aspects of the information environment to attend to most (Ba et al., 2022, Bordalo et al., 2024a).

Asset pricing implications. In the final part of the paper, we include a discussion linking mental models to asset pricing patterns in the cross-section, complementing the evidence on the time-series co-movement between our measure of sentiment and Shiller's CAPE ratio. The associations we document—such as sentiment increasing monotonically with momentum and the time outlook of discussed arguments being substantially more forward-looking for growth firms than value firms—highlight the potential of mental models to deepen our understanding of market expectations, return predictability, and potential mispricing in an interpretable way.

[6] Reflecting these differences, we also show that valuation method usage varies intuitively across industries—for example, DCFs are relatively more common in information and professional services, whereas multiples approaches are more prevalent in retail.

Literature. Our paper contributes to three strands of literature. First, we contribute to the literature in empirical asset pricing that has studied investors' quantitative forecasts, and used them to make progress in our understanding of asset pricing patterns (Greenwood and Shleifer, 2014, Giglio and Kelly, 2018, Giglio et al., 2021, Nagel and Xu, 2022, De La O and Myers, 2021, Bordalo et al., 2019, 2024b, Bastianello, 2024, Ben-David and Chinco, 2024, Decaire and Graham, 2024, and Decaire and Guenzel, 2024). Our paper also contributes to the emerging literature at the intersection of machine learning and finance which often uses text data to study beliefs (Asquith et al., 2005, Giglio et al., 2022, Bybee, 2023, Gormsen and Huber, 2023, Van Binsbergen et al., 2023, Gabaix et al., 2023, Bianchi et al., 2024, Bybee et al., 2024, Charles and Sui, 2024, Chen, 2024, Lopez-Lira and Tang, 2024, Decaire and Graham, 2024, Decaire et al., 2024, Decaire and Guenzel, 2024, Sarkar, 2025), as well as a literature in accounting that has documented variation in the usage of valuation methods and analyzed the accuracy of the resulting price targets (Bradshaw, 2002, Demirakos et al., 2004, Imam et al., 2013, Gleason et al., 2013, Erkilet et al., 2022). We add to this line of research by linking the text data in sell-side analysts' equity reports to their subjective quantitative beliefs. In contemporaneous work, Chen et al. (2025) leverage equity reports to study buy-side analyst recommendations, and De Rosa (2024) and Ke (2025) link sell-side equity reports to IBES forecast data to study memory and attention, respectively.

Second, we contribute to the broader work in finance and economics on measuring mental models and understanding the effect of mental models on beliefs and behavior. A growing body of research uses open-ended and structured survey questions to capture the thought process in people's reasoning in a variety of different fields, including macroeconomics and finance (Bailey et al., 2019, Andre et al., 2022, Chinco et al., 2022, Chopra and Haaland, 2023, Andre et al., 2024a, Andre et al., 2024b, Bauer et al., 2024, Binetti et al., 2024, Stantcheva, 2024, Laudenbach et al., 2024).[7] We contribute to this work by considering a high-stakes field setting.[8] Our approach enables us to link the reasoning and quantitative forecasts of financial professionals. Moreover, unlike these earlier papers, we did not have to elicit information on people's reasoning via additional surveys. This allows us to exploit a whole time-series and

[7] See Haaland et al. (2024) for a recent survey of this literature.

[8] An early mostly theoretical exploration in which analysts use simplified reasoning includes Hong et al. (2007).

cross-section, rather than having to rely on the snapshot in time when the survey was run.

Third, we contribute to an active body of work in behavioral economics that studies how people represent problems, and how they allocate attention (Woodford, 2001, Sims, 2003, Veldkamp, 2011, Gabaix, 2014, Kacperczyk et al., 2016, Gagnon-Bartsch et al., 2018, Gabaix, 2019, Kohlhas and Walther, 2021, Fan et al., 2021, Schwartzstein and Sunderam, 2021, Ba et al., 2022, Bordalo et al., 2023, Bordalo et al., 2024a, Charles and Kendall, 2024, Flynn and Sastry, 2024). Our paper contributes to this literature by measuring both attention and information processing in a field setting.

Organization of the paper. The rest of the paper proceeds as follows. Section 2 introduces a conceptual framework to guide the empirical analysis. Section 3 introduces the data, reviews different text analysis methods, and discusses our LLM-based mental model extraction method. Section 4 presents summarizing facts on the extracted mental model components. Section 5 links mental models to quantitative forecasts, and Section 6 examines the link between mental model heterogeneity and forecast disagreement. Section 7 discusses asset pricing implications, and concludes.

2 Conceptual Framework

The first challenge in analyzing the text of equity analysts' reports is to organize and structure this data in a way that enables a systematic examination. In deciding what data to collect about analysts' reasoning, we start from the canonical present value relation that prices reflect the present value of expected future dividends, and focus on how analysts apply this formula:

$$P_t = \mathbb{E}_t \left[\sum_{t=0}^{\infty} \frac{D_{t+1+j}}{1 + R_{t+1+j}} \right] \tag{1}$$

Applying this formula requires making two high level choices. First, since analysts typically do not forecast an infinite stream of future dividends and discount rates, they must instead choose an appropriate approximation—in other words, they must select a valuation method. A common example is a discounted cash flow (DCF) model, which projects and discounts expected cash flows over a finite horizon, before assuming a constant terminal growth rate. Alternatively, analysts may rely on a more reduced-form, multiples-based approach, inferring value by comparing the firm's metrics (e.g., earnings or sales) to those of comparable peers or its own historical trends. Second, once a particular valuation method is chosen, analysts must decide which features to pay attention to when estimating its relevant inputs. Analyzing the present value relationship in equation (1) (which we can think of as being the most complete valuation method) reveals three key dimensions that analysts may reason through when determining the impact of a given piece of news on valuation: the valuation channel through which it operates (cash flows or discount rates), the time horizon it affects, and its directional impact.

Motivated by this thinking, we define a mental model as a combination of i) a valuation method and ii) a set of attention weights on different inputs, where each input is characterized by a topic, along with its associated valuation channel and time outlook. We discuss our data collection efforts of these features in detail in Section 3. Before turning to data, however, we present a conceptual framework that illustrates how analysts might choose valuation methods and allocate attention across features when faced with limited cognitive resources (Veldkamp, 2011, Gabaix, 2019).

2.1 Setup

To model the core choices analysts make with a minimal working example, we recast the present-value relationship from equation (1) in reduced-form, and model prices as a linear function of different variables:

$$p \;=\; \sum_{k=1}^{K} v_k \, x_k \tag{2}$$

where v_k capture valuation weights, and x_k relate to relevant features that impact valuation. These may include both standard financial metrics (e.g., cash-flow forecasts or discount-rate proxies) as well as harder-to-measure components (e.g., brand strength or regulation).

Given this environment, we introduce the two key elements that analyst ought to choose: valuation methods, and attention allocation across variables. Starting from the first component, we assume that when valuing a firm analysts can choose from a set of valuation methods $m \in M$. Each method m assigns a distinct set of weights $\{m_k\}_{k=1}^{K}$ to the relevant variables $\{x_k\}_{k=1}^K$. Formally, the price target under method m is given by:

$$p^m = \sum_{k=1}^{K} m_k \, x_k, \tag{3}$$

where m_k indicates the weight that method m places on variable x_k.

Definition 1 (Valuation Method)**.** *Let $\{x_k\}_{k=1}^K$ be the relevant input variables for valuing a firm. A* valuation method *m is determined by a vector of weights $\{m_k\}_{k=1}^K$, where each $m_k \in \mathbb{R}$ determines the weight of the corresponding variable x_k in valuation. Under valuation method m the price target is then given by: $p^m = \sum_{k=1}^K m_k x_k$.*

To allow for the role of attention in this framework, we assume that analysts do not observe each feature x_k (e.g., they do not observe future cash-flows or discount-rates precisely). Instead each feature has a prior distribution $x_k \sim N(\mu_k, \tau_{0k}^{-1})$, where μ_k is the prior mean and τ_{0k}^{-1} is the prior variance. Analysts then receive noisy signals:

$$s_k = x_k + u_k, \quad \text{where} \quad u_k \sim N(0, \tau_{sk}^{-1}) \tag{4}$$

where τ_{sk} is the precision of signal s_k. Upon receiving each signal, analysts update their beliefs via Bayesian updating, such that their posterior beliefs are given by: $\mathbb{E}[x_k|s_k] = \mu_k + \frac{\tau_{sk}}{\tau_{sk}+\tau_{0k}}(s_k - \mu_k)$ and $\mathbb{V}[x_k|s_k] = (\tau_{sk} + \tau_{0k})^{-1}$, and the resulting price target is:[9]

$$\mathbb{E}[p^m|\{s_k\}_{k=1}^K] = \sum_{k=1}^{K} m_k \mathbb{E}[x_k|s_k] = \sum_{k=1}^{K} m_k(s - \mu_k)a_k + m_k\mu_k \tag{5}$$

where $a \equiv \tau_{sk}/(\tau_{sk} + \tau_{0k})$. Finally, we allow analysts to choose the precision of each signal (τ_{sk}) by allocating attention across variables, subject to a linear budget constraint: $\sum_{k=1}^K c_k \tau_{sk} \leq C$, where c_k denotes the marginal cost of devoting one more unit of increasing attention to feature x_k, and C is the total attention budget.

Definition 2 (Attention)**.** *Analysts* attention *is captured by the precision parameters $\{\tau_{sk}\}_{k=1}^K$ of the signals they acquire about each feature $\{x_k\}_{k=1}^K$. Allocating more attention to a particular features x_k increases τ_{sk}, resulting in a more precise signal for that feature.*

[9] We assume that all features x_k are uncorrelated with one another, so that signal s_k is only relevant for forecasting feature x_k.

Having specified how both valuation methods and attention weights enter our model, we now define a mental model as combination of these two components.

Definition 3 (Mental Model)**.** *A mental model $\mathcal{M}(m, \tau_s)$ is defined by the combination of (i) a valuation method $m = \{m_k\}_{k=1}^K$, and (ii) a vector of attention weights $\tau_s = \{\tau_{sk}\}_{k=1}^K$.*

Given this framework, we are interested in understanding how analysts should optimally choose a valuation method and allocate attention across features. To do so, we first establish a benchmark in which analysts choose attention weights under the true valuation function (i.e., by setting $\{m_k\}_{k=1}^K = \{v_k\}_{k=1}^K$). We then consider how the adoption of alternative valuation methods leads to deviations from this benchmark.

2.2 Attention Allocation Across Variables Under the "True" Valuation Model

To gain intuition in the simplest possible setting, we focus on the case with just two features. Moreover, in this section we assume that analysts use the true valuation function as their valuation method, i.e., $\{m_k\}_{k=1}^K = \{v_k\}_{k=1}^K$. Analysts then choose how much attention to allocate to each feature by minimizing mean squared forecast errors, subject to their linear budget constraint on attention:[10]

$$\min_{\tau_{s1}, \tau_{s2}} \quad \mathbb{E}\left[(p - \mathbb{E}[p|s_1, s_2])^2\right] \qquad s.t. \qquad c_1\tau_{s1} + c_2\tau_{s2} \leq C, \quad \tau_{s1} \geq 0, \quad \tau_{s2} \geq 0 \tag{6}$$

Since analysts in our model perform rational Bayesian updating and use the true valuation function for valuation, their forecasts are unbiased on average. Therefore, the problem is

[10] Analysts have incentives to focus their efforts on producing informative valuation theses and credible price targets and recommendation. Two major organizations evaluate and rank analyst performance: Institutional Investor (II) and the Wall Street Journal (WSJ). While each emphasizes different aspects of sell-side analyst activity, the criteria they use are complementary. Institutional Investor conducts surveys of key market participants—including portfolio managers and buy-side analysts—asking which analysts provided the most valuable service and advice. This holistic assessment tends to reward analysts for high-quality valuation frameworks and the ability to identify and frame the firm-specific factors most relevant for understanding price dynamics. In contrast, the Wall Street Journal ranking focuses narrowly on the performance of analysts' stock recommendations (i.e., buy/hold/sell) relative to realized returns.

equivalent to minimizing their posterior variance:

$$\min_{\tau_{s1},\tau_{s2}} \quad \mathbb{V}[p|s_1, s_2] = v_1^2 \left(\tau_{01} + \tau_{s1}\right)^{-1} + v_2^2 \left(\tau_{02} + \tau_{s2}\right)^{-1} \tag{7}$$

$$s.t. \quad c_1\tau_{s1} + c_2\tau_{s2} \leq C, \quad \tau_{s1} \geq 0, \quad \tau_{s2} \geq 0 \tag{8}$$

Setting up the Lagrangian and taking first order conditions:

$$v_1^2(\tau_{01} + \tau_{s1})^{-2} = \lambda c_1 \tag{9}$$

$$v_2^2(\tau_{02} + \tau_{s2})^{-2} = \lambda c_2 \tag{10}$$

which equates the marginal benefit of acquiring an additional unit of information with its marginal cost. From this simple expression, it is clear that the marginal benefit of acquiring information about x_k is increasing in v_k (its relevance for valuation), and decreasing in τ_{0k} (its volatility).

Proposition 1 (Allocation of Attention Across Features)**.** *Ceteris paribus, an analyst will allocate more attention to feature x_k when it has a greater weight in the price function (higher v_k), is more volatile (lower τ_{0k}), and is less costly to acquire information about (lower c_k).*

Proof. All proofs are in Appendix A, unless otherwise stated. □

These comparative statics help us understand why analysts may allocate differential attention to features both in the cross-section and over time. Proposition 1 implies that if two industries or stocks differ in how relevant, volatile or costly different variables are (v_k, τ_{0k}, c_k), then analysts will allocate attention differently when valuing them. For example, brand perception may be highly relevant in consumer-goods or tech-platform firms, whereas regulated utilities may demand more focus on policy changes or regulatory filings. Similarly, pharmaceutical companies often exhibit greater volatility in research outcomes, prompting analysts to devote more resources to assessing clinical data than they would for stable consumer-staples companies. These comparative statics also apply in the time-series: if any of the relevant parameters vary over time (e.g., if a feature becomes more volatile), then analysts' attention allocation will also vary.[11] Combining the two first order conditions and using the budget constraint we then obtain:

$$\tau_{s1}^{*}=\frac{C+c_2\tau_{02}-\frac{v_2}{v_1}\sqrt{c_1}\sqrt{c_2}\tau_{01}}{c_1+\frac{v_2}{v_1}\sqrt{c_1}\sqrt{c_2}}. \tag{11}$$

and $\tau_{s2}^{*}={}^{(C-c_1\tau_{s1}^{*})}/_{c_2}$. If either of the interior solutions is negative, analysts optimally set the corresponding precision to zero, i.e., they pay no attention to that feature, and allocate the full attention budget to the remaining variable.

Proposition 2 (Sparsity)**.** *If the marginal benefit from acquiring the first unit of information is lower than its marginal cost ($v_k^2(\tau_{0k})^{-2}<\lambda c_k$), then analysts will pay no attention to that variable, setting $\tau_{sk}=0$ and $\tau_{sj}={}^{C}/_{c_j}$. A variable is more likely to be neglected if it is less relevant for valuation (low v_k), more stable over time (high τ_{0k}) or has a higher marginal cost (high c_k).*

Thus, analysts allocate attention only to variables with a sufficiently high marginal value of information. Variables that are less important for valuation, highly stable, or costly to learn about are more likely to be ignored, leading to sparse mental representations.

2.3 Endogenous Choice of Valuation Method

Thus far, we have assumed that analysts use the true valuation function when valuing firms (i.e., we imposed that $\{m_k\}_{k=1}^{K}=\{v_k\}_{k=1}^{K}$). In practice, however, analysts often have access to a range of valuation methods that make simplifying assumptions and systematically assign more weight to some variables over others. For example, a DCF model tends to emphasize projected cash-flows and discount rates, whereas a multiples-based approach may lean more on peer performance or recent trends. In this section, we analyze how analysts should choose among (potentially misspecified) valuation methods, and how these choices affect their allocation of attention across variables.

Specifically, we assume that analysts know the true valuation model in (2), and that they

[11] Moreover, notice how the expression in (11) shows how attention across all features may change even when the volatility of a single variable changes.

have access to a range of different valuation methods $m = \{m_k\}_{k=1}^{K}$.[12] Analysts then choose $\mathcal{M}(m, \tau_s)$ to minimize the mean squared forecast error:

$$\mathbb{E}\left[(p - \mathbb{E}[p^m|s_1, s_2])^2\right] = \mathbb{V}\left[p - \mathbb{E}[p^m|s_1, s_2]\right] + \mathbb{E}\left[p - \mathbb{E}[p^m|s_1, s_2]\right]^2 \tag{12}$$

$$= \sum_{k=1,2}\left(\frac{v_k^2 - (v_k - m_k)^2}{\tau_{0k} + \tau_{sk}} + \frac{(v_j - m_j)^2}{\tau_{0j}}\right) + \left(\sum_{j=1,2}(v_j - m_j)\mu_j\right)^2 \tag{13}$$

When analysts use the true model ($m_k = v_k$), this expression reduces to the objective function in (7). Instead, adopting an alternative mental model ($m_k \neq v_k$) leads to changes in both the variance and bias components of the total mean squared error.

To solve the model, we proceed in two steps. First, we compute the optimal attention allocation τ_s for a given valuation method $m = \{m_k\}_{k=1}^{K}$. Second, we solve for the valuation method that minimizes mean squared errors under these optimal attention weights. Starting from the first step, we notice that attention weights τ_{sk} appear only in the first term of the objective function in (13). Therefore, analysts choose their attention allocation to minimize:

$$\min_{\tau_{s1}, \tau_{s2}} \frac{v_1^2 - (v_1 - m_1)^2}{\tau_{01} + \tau_{s1}} + \frac{v_2^2 - (v_2 - m_2)^2}{\tau_{02} + \tau_{s2}} \qquad s.t. \qquad c_1\tau_{s1} + c_2\tau_{s2} = C \tag{14}$$

Relative to the objective function in equation (7), this expression shows that analysts are now more cautious about allocating attention to variables for which the true and misspecified models diverge more. This becomes even clearer when solving for the resulting optimal attention allocation, and comparing it to the true valuation model benchmark in (11):

$$\tau_{s1}^m = \frac{C + c_2\tau_{02} - \sqrt{\frac{v_2^2 - (v_2 - m_2)^2}{v_1^2 - (v_1 - m_1)^2}}\sqrt{c_1}\sqrt{c_2}\tau_{01}}{c_1 + \sqrt{\frac{v_2^2 - (v_2 - m_2)^2}{v_1^2 - (v_1 - m_1)^2}}\sqrt{c_1}\sqrt{c_2}}. \tag{15}$$

Proposition 3 (Attention Allocation when Using a Misspecified Valuation Method)**.** *When analysts use a valuation method m that deviates from the true valuation function v (i.e., $m_k \neq v_k$ for some k), they allocate more attention to variable x_k and less attention to variable x_j—relative to the benchmark with no deviations—if and only if $\left(\frac{v_k - m_k}{v_k}\right)^2 < \left(\frac{v_j - m_j}{v_j}\right)^2$. That*

[12] The fact that analysts know the true valuation model simply reflects the fact that analysts often have a sense of the relevant importance of different variables in valuing a firm.

is, analysts allocate relatively more attention to variables where the true and misspecified models are more closely aligned, and relatively less attention to variables for which the discrepancy is larger.

For instance, consider a growth stock whose true model places substantial weight on long-term growth prospects. A DCF method—which explicitly projects future cash flows—aligns more closely with this emphasis than a multiples-based approach that focuses on peer comparisons or historical trends. According to the proposition, because the DCF method's discrepancy from the true model is smaller in the growth dimension, analysts using DCF will allocate more attention to signals that inform these prospects, whereas those using a multiples-based approach will allocate less attention to them.

Moving to the second step, analysts choose the valuation model (and corresponding attention weights) that minimizes the expected mean squared error. Substituting the optimal attention weights in (15) into the mean squared errors expression in (13), we find that the resulting MSE from using valuation model m can be written as:

$$MSE(m) = \underbrace{\frac{\left(\sum_{k=1,2}\sqrt{c_k\left(v_k^2-(v_k-m_k)^2\right)}\right)^2}{C+c_1\tau_{01}+c_2\tau_{02}} + \sum_{k=1,2}\frac{(v_k-m_k)^2}{\tau_{0k}}}_{\text{variance}} + \underbrace{\left(\sum_{k=1,2}(v_k-m_k)\mu_k\right)^2}_{\text{bias}} \tag{16}$$

This equation shows that using a misspecified model ($v_k \neq m_k$) unambiguously increases the bias term, but it may decrease the variance term. Consequently, analysts face a bias/variance trade-off, meaning that it is not necessarily optimal to choose the correct valuation model. Instead, accepting some degree of model misspecification (and thus some bias) can reduce mean squared error if it allows analysts to reallocate attention more effectively toward variables that are more important, more volatile, and are less costly to acquire information about.

Proposition 4 (Optimal Choice of Valuation Method)**.** *When analysts can choose among multiple valuation methods, they may optimally deviate from the true valuation model, as they face a bias-variance trade-off. Relative to a baseline with no wedges, it can be optimal to introduce a wedge $w_k^2 = (v_k^2 - m_k^2) > 0$ on variable x_k if doing so reduces overall MSE.*

Ceteris paribus, this becomes more likely when x_k is less relevant (lower v_k, μ_k), when the value of learning about the other variable is higher (lower τ_{0j}, c_j, or higher v_j), or when the attention budget C is lower. Conversely, analysts are more likely to align their model with the true valuation function (i.e, set $w_k^2 = 0$) when x_k is more relevant, the value of learning about the other variable is lower, or the total budget on attention is higher.

Returning to our earlier example, growth stocks place a high weight on long-run outcomes. Therefore, when valuing growth stocks, analysts are more likely to use a DCF model that minimizes bias on that crucial dimension. In contrast, with value stocks, long-run growth is less central, making analysts more comfortable using simpler multiples–even if doing so might introduce some bias along that dimension (which is now less important).

This analysis illustrates how the choice of valuation methods both shapes and is shaped by what analysts attend to. These considerations can explain why analysts may adopt different valuation methods and attention weights across industries or firm types or over time, as we examine empirically below.

2.4 Empirical Predictions

There are four sets of predictions we test empirically. First, analysts attend to variables that are more relevant, more volatile, and are less costly to acquire information about. This should hold both in the time-series and in the cross-section. Second, the choice of valuation methods and attention are tightly interlinked. When analysts agree on valuation methods, they should attend to more similar information. Conversely, when pairs of analysts covering the same firm at the same point in time disagree on valuation methods, they will also attend to very different information, leads to different mental representations. Third, analysts choose valuation methods based on the relevance of different variables, and seek to minimize discrepancies with the true valuation function along dimensions that matter the most. Finally, disagreement may be due to either differences in attention or differences in valuation weights.

3 Data Collection and LLM-Based Mental Model Extraction

In this section we describe our data collection method for measuring analysts' mental models. The first step required collecting equity analyst reports from Refinitiv. The second step entailed extracting information regarding analysts' quantitative forecasts, their valuation methods, and their associated reasoning.

3.1 Equity Analyst Reports – Data Collection

We first downloaded the universe of equity analyst reports available on Refinitiv, and then constructed a subsample to use for the more cost-intensive parts of our analysis.

Universe of equity reports in Refinitiv. We download the universe of analyst equity reports from Refinitiv Eikon (LSEG Workspace) between 2000 and 2024, focusing on the 43 most common brokerage houses in the data following Decaire and Graham (2024). We apply a set of light data filters during the extraction process, including a maximum page count of 30, a restriction to English-language and company-specific reports (excluding industry or macro reports), and excluding reports related to M&A activity. This results in a total dataset of 2.1 million reports, of which we use the near-universe available at the time of the analysis—1.6 million reports—prior to the final batch of downloads, covering more than 11,000 firms worldwide.

From these reports, we instruct Gemini to extract the price target estimates, together with analyst names, locations, and phone numbers. Finally, we process each downloaded report—available in PDF format—using a set of Python routines and LLM prompts to extract the underlying text,[13] and we collect information on i) all valuation methods mentioned in a given report, ii) the "main" valuation method used by the analyst (i.e., the valuation method tied to the price target estimate), and iii) if multiples were used, whether it was obtained through DCF or forward looking considerations, historical comparisons, and peer or industry comparisons (see Appendix C for details on the prompt). This data step allows us to shed light on how analysts' mental model are reinforced by, and shape valuation method choices.

[13] For text extraction, we use Google's Gemini as well as Meta's Llama-2-70.

Subsample used for main analysis. As discussed in detail further below, the core of our analysis requires, in addition, the extraction of every line of reasoning an analyst mentions to justify their price target. We collect this as a list of topics, together with their associated valuation channel, time-outlook, sentiment, and a snippet that allows us to locate the argument in the report. This process is very computationally intensive, and it is cost-prohibitive to use all 1.6 million reports. We therefore construct a subsample of reports using a two-pronged approach.

First, we start from the 81,757 reports previously used in Decaire and Graham (2024), which were already partially processed for text and price target information, but not for mental model extraction. We refer the reader to Decaire and Graham (2024) for details but briefly, these reports were downloaded restricting the time window to reports published in the first quarter of the calendar year (January 1 to April 1) from 2000 to 2023, and by focusing on DCF-based reports. Decaire and Graham (2024) show that the resulting dataset is comprehensive, representative, and comparable to commercial datasets such as IBES—across dimensions including industry coverage, geographic distribution, and brokerage house representation. Second, since we are also interested in studying variation in valuation methods across analysts, we randomly select an additional 30,000 reports from the subset of the full sample that meets the time window restrictions used in Decaire and Graham (2024), as discussed above. We then follow Decaire and Graham (2024) in removing randomly selected reports with fewer than 200 words, yielding a final main sample of 111,270 reports, covering 11,308 firms and more than 6,400 analysts. Given our two-pronged approach, we include analyses both across and within valuation methods.

3.2 LLM-Based Text Analysis for Topic Interpretability and Scalability

The existing finance and economics literature has relied on two main approaches to extract structured information from financial text—such as equity reports, news articles, and surveys—namely: (i) standard machine learning techniques, particularly LDA models (Decaire and Guenzel, 2024; Bybee et al., 2020), and (ii) human workers (e.g., Andre et al., 2024a). Both of these approaches have benefits and limitations. Employing LLMs, such as Claude from Anthropic (our chosen model for the extraction of detailed mental models, see Section 3.3), overcomes the respective limitations of these earlier methodologies, while enhancing their key strengths.

LDA models efficiently identify topics in large document collections; however, the resulting topics are often difficult to interpret, limiting the types of possible research questions. In the context of equity reports, LDA tends to produce broad thematic categories—grouping content by industry, institution, or region. For example, Decaire and Guenzel (2024) use LDA to extract topics from equity reports and document the associated keywords. While this approach is well-suited for certain questions (such as those in Decaire and Guenzel (2024)), it is ineffective when research requires *intuitive* and *interpretable* topics or clear causal reasoning—both of which are crucial for extracting forecasters' mental models.

In contrast, human workers can interpret text and extract topics that are both intuitive and contextually meaningful. Human judgment is also well suited for identifying causal relationships in text, broadening the range of possible research questions (e.g., Andre et al., 2024a). However, human annotation limits scalability, typically restricting analyses to a few hundred documents. Consistency is also a concern—different workers may identify different sets of topics, even when given the same task, particularly in long and complex documents. Implementing cross-checks further restricts scalability, making large-scale text analysis impractical.

LLMs enable us to generate clear and interpretable topics, comparable to those produced by human workers, while also allowing to process a much large number of reports—facilitating more comprehensive analysis and enabling deeper exploration of important heterogeneities. In addition, LLMs support repeated prompting at relatively low cost, allowing for robustness and validity checks similar to those feasible with other ML methods such as LDA, but not with human workers.

3.3 LLM-Based Extraction of Mental Models

3.3.1 Employed Approach

The purpose of our LLM approach is extract all valuation narratives from equity reports—that is, all discussed topics (e.g., market share, M&A, innovation, supply chain) along with their associated valuation channel (e.g., impacting sales, costs, or margins), time outlook, and sentiment. Our data collection process is carefully designed to allow for the possibility that multiple arguments may reference the same topic in different ways. To extract the mental model from the text of equity reports, Specifically, we employ Claude 3.5 Sonnet from Anthropic and introduce a new, three-step extraction method.[14]

Steps 1 and 2. In the first two steps, our objective is to extract three categories of information from equity reports.

First, we task the AI with identifying all arguments in each report that relate to the topics used to explain and justify the analyst's valuation forecast. We deliberately impose *no* ex-ante structure on the type of topics the AI is allowed to extract, nor do we constrain it to a predefined set of labels. (We will introduce additional structure on topics in Step 3, as discussed below.) Besides topic extraction, the AI classifies each extracted argument along two additional attributes: sentiment (positive, neutral, negative, or unclear) and time outlook (past, present, near future (1–3 years), distant future (3+ years), and unclear). To support this classification, the report's publication date is provided.

Second, we task the AI to classify each argument as firm-related, industry-related, or macroeconomic in scope. It also determines the associated valuation channel, identifying whether the topic pertains to revenues or sales (top-line), costs, profitability margins, bottom-line outcomes (e.g., EBITDA, cash flows, earnings), discount rates, or relative valuation statements (e.g., comparisons to peers or historical averages). These dimensions further allow us to reconstruct the structure of the analyst's mental model as expressed in the report.

Third, to facilitate output verification, we extract two sets of words directly from the report. For one, we ask the AI to identify one or two keywords that best characterize the sentiment associated with each topic. Additionally, we extract the first five words linked to the relevant section of the text, enabling us to locate the topic within the document and assess whether the extracted topic is contextually plausible.

The key distinction between Steps 1 and 2 lies in the text we provide to the AI. In Step 1, we divide each report into 200-word segments and prompt the AI iteratively on each segment.

[14] Appendix E provides an example excerpt from an equity report along with the full corresponding model output.

The 200-word threshold is motivated by the fact that the shortest documents in our sample contain at least 200 words, and we provide further discussion on this in Appendix B. Crucially, this *chunking method*—i.e., segmenting reports into smaller "sub-reports"—improves the AI's ability to detect the full set of topics and enhances output consistency, as we detail in Section 3.3.3.

Step 2 addresses a potential concern with the chunking method—namely, that segmenting the text introduces the risk of missing topics that span multiple segments and are therefore not fully captured. In Step 2, we provide the AI with both the full equity report and all all topics identified in the first step. The AI then re-evaluates the report holistically to identify any potentially missing topics, and adds these topics to the final output if necessary.

Step 3. In Step 3, we group the unstructured topics extracted in the first two steps into 139 standardized labels. The construction of this label set is guided by two inputs: a word2vec-based clustering of 200 commonly extracted topics, and financial intuition to ensure that the resulting labels are intuitive and economically meaningful in the context of valuation (see Appendix B for further details). We use a two-tier structure for the topic labeling. We first instruct the AI to assign each raw topic to one of 128 refined standardized labels (Appendix Table B1). If no clear match is found, the AI selects from a second, coarser set of 11 categories (Appendix Table B2). This approach reflects the way analysts structure discussion in equity reports. For example, an analyst might explicitly state, "Inflation pressure will increase the cost of goods sold," linking inflation to costs. In other cases, they may simply write, "The cost of goods sold will increase," omitting the causal factor. To capture both types of expressions systematically, our final prompt is designed to favor more granular labeling when possible, while remaining flexible enough to include broader statements when necessary.

3.3.2 Basic Report Anatomy: Location of Topics, Sentiment, and Time Outlook

Before comparing our multi-step prompt to the performance of a naïve, single-step prompt, we first provide detail on the basic "report anatomy" of mental models. For example, where in the report does the AI tend to collect information—both overall and with respect to specific dimensions, such as firm- versus macro-level topics or past- versus future-oriented outlooks? This evidence is useful context for the comparison with the naïve prompt, as it provides insight into what is likely to be omitted by less sophisticated prompting strategies.

Consistent with the intuition that introductions and conclusions of reports are more content-dense, we find that the AI extracts slightly more information and topics from the very beginning and end of reports (Panel A of Figure I; see also Panel C of Figure II, discussed further below). At the same time, information is consistently extracted throughout the full length of reports.

Moreover, the nature of arguments varies systematically across a report's progression (Panels B to D of Figure I). Early sections ("executive summary") tend to be more backward-looking and review past developments and earnings. The middle section ("investment thesis") is more forward-looking, with sales-related topics dominating the discussion. This is consistent with our findings below on the dominance of top-line items in analyst narratives (Section 4.2), as well as survey evidence in Graham (2022), which documents that managers primarily focus on forecasting sales. In the final section ("key risks"), analysts tend to focus on future downside scenarios, with sentiment declining sharply and attention shifting toward industry- and macro-level topics.

One important takeaway from this evidence on report anatomy is that different types of information appear in different parts of analyst reports. As a result, less sophisticated prompting strategies that disproportionately focus on, for example, the beginning of a report—rather than systematically processing the full text, as ensured by our chunking approach—risk producing mental models that are not only incomplete but also systematically distorted.

3.3.3 Diagnostic Tools: The Importance of a Multi-Step LLM Approach

This section provides direct evidence on the importance of using a sophisticated LLM prompting strategy to extract detailed and comprehensive narratives and mental models from analyst reports—at least given the current capabilities of LLMs.

Comprehensiveness of Extracted Topics. We first compare the comprehensiveness of our extracted mental models to those produced by a single-step naïve prompt—that is, a prompt that collapses the three-step procedure from the previous section into a single query. We evaluate comprehensiveness across the distribution of report lengths. The modal report contains just under 1,000 words, but the distribution is right-skewed, with a substantial number of longer reports (Panel A of Figure II).

Panel B of Figure II shows that the output generated by a naïve one-step prompting strategy exhibits both lower breadth—fewer topics collected—and lower depth—fewer individual arguments per topic—relative to our multi-step approach, where an argument is a combination of topic, valuation channel, time outlook, and associated sentiment. When provided with the full text at once, the AI tends to extract only one argument per topic, a pattern that holds across reports of varying lengths. By contrast, our multi-step prompt consistently captures multiple arguments per topic, enabling a more comprehensive reconstruction of analyst narratives and, for example, a more accurate measure of narrative (dis)agreement across analysts later in Section 5.[15] Moreover, the AI displays a form of "laziness" under the one-step strategy, concentrating data collection almost exclusively at the beginning of the report (Panel D), in contrast to our multi-step approach, which ensures coverage throughout the full document (Panel C; see also Panel A of Figure I). This pattern is especially concerning given that, as discussed above, the nature of information varies systematically across different sections of the report—in terms of topics and valuation channels, sentiment, as well time outlook.

As one concrete example corroborating the importance of collecting individual arguments and narratives *throughout* the report, Appendix Table E.2 shows that narrative sentiment at all points in the report significantly predicts analysts' price target forecasts—and does so positively, as one would expect. This result holds both when sentiment from different sections is included separately and when it is included jointly as predictors.[16] In light of this, prompting strategies that fail to capture the full set of narratives across the report will be less effective in uncovering the underlying processes that shape analysts' quantitative beliefs—and the heterogeneity in those processes across forecasters.

[15] Appendix Table E.1 confirms this pattern using within-report statistics. In the average report, 13% of topics are linked to more than one valuation channel (e.g., sales, costs), and in over 23% of cases, a topic is associated with multiple sentiments (e.g., inflation is expected to *positively* impact valuation through sales but *negatively* through costs).

[16] The only exception is that sentiment at the very end of an report does not tend to provide incremental predictive power once sentiment from rest of the report included.

Stability of Extracted Topics. Recent advances in LLM performance and the sharp decline in associated costs also allow us to introduce a novel method for assessing the robustness of extracted mental models and quantifying the degree of AI-induced variability, in an approach that is akin to evaluating coefficient stability across bootstrapped samples, or to asking multiple equally well-trained human workers to extract mental models from the same report.

Specifically, we prompt the AI to extract the mental model from a given equity report *repeatedly* (ten times) and compute the Jaccard topic similarity score for all pairs of attempts, defined as:

$$\text{Jaccard similarity} = \frac{|A \cap B|}{|A \cup B|} \tag{17}$$

where A and B represent the sets of topics extracted in attempts A and B of a given report, $|A \cap B|$ is the number of overlapping topics across attempts, and $|A \cup B|$ is the total number of distinct topics identified in at least one of the attempts. The Jaccard similarity thus quantifies the degree overlap between the topic sets, ranging from 0 (no overlap) to 1 (complete overlap). We implement repeated prompting for both our multi-step approach and the naïve single-step approach, using a stratified random subsample of 240 reports.[17]

With our multi-step approach, the Jaccard score for topic overlap across repeated prompts is consistently very close to 1, across the full distribution of report lengths. This confirms that the LLM output is highly stable across repeated LLM extractions (Panel E of Figure II). By contrast, the naïve method exhibits greater variability in topic output, with Jaccard similarity scores consistently lower than those produced by the multi-step prompt. These discrepancies are even more pronounced when we evaluate Jaccard similarity as a function of the number of arguments in Panel F. Our multi-step prompt continues to yield stable outputs, reflected in consistently high Jaccard scores across number of arguments. The single-step prompt, by contrast, performs substantially worse: as the number of extracted arguments exceeds 10, Jaccard topic similarity drops below 0.5, indicating that once the naïve prompt moves beyond a few arguments, the topics it selects become highly inconsistent across iterations. In

[17] We stratify the random subsample based on report length, by increments of 100 words and randomly select 5 reports in each stratum. This allows us to study the robustness of the LLM extraction approach over the entire range of report lengths using a well-populated subsample.

other words, beyond the initial set, the additional arguments and topics it generates appear largely arbitrary.[18]

3.4 Descriptive Versus Predictive Information Processing Using LLMs

Before turning to our main results in Sections 4 to 7.2—on summary facts about analyst mental models, their associations with quantitative forecasts, and their implications for asset pricing—we highlight one final aspect of our LLM extraction procedure: its *descriptive* nature.

LLMs can be used for a range of tasks involving unstructured text, which broadly fall into two categories: (i) descriptive extraction and (ii) predictive generation. Our approach is descriptive—we ask the LLM to summarize and structure information present in the report, rather than to predict outcomes based on that information. To ensure this, we set the temperature and top_p parameters to zero, which limits variation and prevents the model from introducing content beyond what is explicitly present in the input. To further constrain the model to rely only on the provided text, we require it to validate each output by supplying direct excerpts from the report (cf. the example in Appendix E). This ensures that extracted topics, sentiment, and time outlooks are grounded in the original document rather than inferred from the model's pre-training.

These design choices are important given concerns about potential training data leakage in LLMs. Sarkar and Vafa (2024) show that pre-training can induce look-ahead bias in predictive tasks, with models at times relying on training data rather than the provided input. Had we employed a predictive rather than fully descriptive prompt, such leakage could blur the line between what the analyst wrote and what the model already "knows." That said, recent evidence suggests that leakage-induced biases may be modest in certain financial contexts (Engelberg et al., 2025; He et al., 2025).

Nonetheless, to go one step further in addressing any potential concerns about leakage, we conduct an additional empirical validation exercise. We find that 100% of the 5-word text

[18] The "missing" dots for the single-step prompt beyond just above 20 arguments reflect the fact that it never extracts more than this number of arguments to begin with (cf. Panel B of Figure II)—let alone consistently across repeated extraction attempts, as shown in Panel F.

snippets used to validate topic extraction are indeed drawn from the original documents.[19] This provides evidence that the extracted mental model components do indeed reflect the report content, not the AI's prior knowledge.

4 Mental Model Summary Facts

This section presents aggregate patterns and features of both components of analysts' mental models. We start by describing key properties of the valuation methods they use, and then provide evidence on average narrative focus, time outlook, and sentiment, along with initial evidence on heterogeneity over time and across industries. We end by corroborating our theoretical framework's prediction that the choice of valuation methods and attention weights are tightly interlinked.

4.1 Valuation Methods

We begin by presenting statistics on the basic properties of valuation methods, illustrated in Figure III. A key motivation for extracting valuation methods from the data is that these valuation choices can both reflect and reinforce the structure of analysts' mental models, as outlined in our conceptual framework, and as we further discussed in Section 5.

Panel A of Figure III shows that across the near-universe of equity reports (for which the LLM extraction identified the valuation method(s) used), approximately 35% rely on discounted cash flow (DCF) methods, a nearly equal share use price-to-earnings (P/E) multiples, and about 65% use at least one multiples-based approach. These statistics account for the fact that a single report often includes more than one valuation technique, with 30%

[19] In rare cases (1.4% of snippets), the AI fixed typos or combined non-contiguous text to enhance interpretability. For instance, in a January 13, 2011 report for RAND.AS from RBC Capital Markets, the AI extracted "*expect further recovery from later-cycle*" to represent the text "*expect further recovery from* ***[the]*** *later-cycle*." We view these instances as acceptable in context.

of reports using a blend of different methods on average.[20]

Panel C of Figure III sheds light on the time trends underlying these averages. The use of DCF methods has increased relatively steadily over time, with approximately 40% of reports now relying on them, while the use of P/E multiples has declined from around 50% in 2004 to roughly 30% in recent years. Other approaches—such as blended models and sum-of-the-parts valuations—have also seen a slight increase in popularity over time.

Additionally, Panel D of Figure III reveals substantial variation in valuation methods across industries (based on the ten most frequent industries in the sample). DCF methods are most common in utilities, information, and professional services, and least common in manufacturing, construction, and finance. Multiples (as well as P/E multiples specifically) are most frequently used in retail and manufacturing, and are rarely applied in real estate and mining. The latter two industries more often rely on net asset value. Blended models and sum-of-the-parts approaches are especially prevalent in the utilities sector.

The heterogeneity across industries is consistent with the theoretical framework in Section 2, which posits that analysts select among available valuation methods to minimize discrepancies with the true model along the most economically relevant dimensions. The relatively higher usage of DCF methods—structured to emphasize future cash flows—in information and professional services is consistent with the forward-looking nature of these industries, which often focus on innovation and long-run investment horizons. Similarly, the dominance of multiples-based approaches in industries such as retail is consistent with their emphasis on near-term performance and observable comparables. Figure IV presents further corroborating evidence consistent with the cross-sectional theoretical predictions: DCF methods are, relatively speaking, more commonly used for small, young, and growth

[20] While much of the analysis that follows on how valuation methods relate to quantitative forecasts, disagreement, and errors focuses on contrasting DCF and multiples-based approaches, Panel B of Figure III highlights additional variation within multiples-based methods that we have collected. Conditional on the use of at least one multiples-based valuation, firm-based reference points (represented by the union of the three left-most bars) are more common than peer-based reference points (rightmost bar). These shares can exceed 100%, as reports often include more than one approach. The middle bars (fourth to seventh bar) illustrate a subset of cases contributing to the firm-based reference points shown in the first three bars. Additionally, historical reference points are less common than current or forward-looking ones. We caveat these statistics noting that the LLM-based extraction may somewhat overstate the use of future reference points when it classifies a trailing multiple applied to a future estimate as future-oriented. Appendix C and Appendix F provide the prompt we used, together with several examples of reference point cases from our sample.

firms, while multiples-based approaches dominate in larger, more mature, and value firms.

4.2 Attention Allocation

The first fact that we establish is that mental models are topically sparse yet nuanced. Table I shows basic summary statistics of the lines of reasoning we extract, using the main subsample of 111,270 reports described in Section 3.1. The average report contains 43 distinct lines of reasoning, comprising only 18 out of 139 distinct topics per report, with each topic being discussed through 2.4 individual arguments.[21,22]

At an aggregate level, Figure V shows that analysts predominantly focus on top-line items, firm-related topics, and forward-looking arguments. Moreover, attention to top-line items and firm-related topics is pro-cyclical, while focus on discount rates and macro-related topics is counter-cyclical. Specifically, Panel A of Figure V and Panel B of Table I show that when we combine topics into valuation channels sales account for the biggest fraction of discussed topics (40–50%), followed by earnings/cash flows (20–30%), costs (15%), profitability margins (10%), and discount rates (5–10%). Panel B shows that analysts' discussions predominantly center on the forecasted firm itself (75–80%), rather than on macroeconomic (5-10%) or industry-level (15–20%) considerations. Panels C and D of Figure V show that, on average, most arguments in analysts' reports are forward-looking, with near-future discussions more common than those focused on the distant future.

Finally, Panels E and F of Figure V plots average narrative sentiment over time (coded as $+1$ = positive, 0 neutral, and -1 = negative), and shows how it comoves closely with Shiller's CAPE ratio ($\rho = 0.84$). When we examine which features are associated with more and less positive sentiment, Appendix Figure E.10 shows that sentiment is most volatile for arguments that relate to the near-future, while distant-future sentiment is both more stable and more positive than sentiment at other horizons.

At a finer level, Figure VI presents a heatmap showing the average share of attention that

[21] An argument is defined as a unique combination of topic, valuation channel, time outlook, and sentiment. For example, analysts may talk about inflation (a topic) both in terms of costs and in terms of sales (two different valuation channels), resulting into two different lines of reasoning associated with the same topic.

[22] Unsurprisingly—but reassuringly, from the perspective of comprehensively extracting mental models (cf. Section 3.3.3)—longer reports contain a broader set of topics and arguments (Panel B of Figure II).

the typical report in our sample allocates to each topic in a given year.[23] Topics vary both in the amount of attention they receive, and in how stable their attention allocation is. Topics related to fundamental drivers of cash flow generation receive a large and persistent share of attention. These include, for example, customer demand, market expansion and market share, capital expenditures, cost structure and efficiency, and mergers and acquisitions. In contrast, some topics receive heightened attention only during specific periods. Most notably, topics related to public health (e.g., pandemics) and supply chain issues become prominent starting in 2020, taxation shows up during the first term of the Trump administration, and inflation becomes prominent in the latter part of the sample.

Appendix Figures E.1 through E.8 show how attention to topics varies across industries. For example, in resource-dependent sectors such as oil & gas and mining, forecasters tend to focus on commodity and raw material prices, oil and gas prices, and production capacity—firms' ability to secure natural resource reserves (Appendix Figure E.5). By contrast, dynamic and innovation-driven industries, such as the information sector and professional and scientific services, place greater emphasis on capital investment, customer acquisition, market share, and product and service mix (Appendix Figures E.3 and E.4). Reflecting—and possibly reinforcing—these differences in topic attention, these cross-industry patterns are also strongly correlated with the variation in valuation method usage described above.

Finally, we can look at the interaction of coarser and finer categories by looking at which topics are associated with higher or lower sentiment and time outlook. Appendix Figures E.11 and E.12 show intuitive relationships. For example, among the most forward-looking topics are broad economic trends such as demographic change and the energy transition, as well as certain firm-specific themes like management guidance. In contrast, the most backward-looking topics include natural disasters and employee safety, suggesting that analysts address these risks reactively rather than proactively. Similarly, topics associated with the most positive sentiment include automation and market expansion, while recessions and wars are almost exclusively discussed with negative sentiment.

These patterns further highlight the advantages of working with directly interpretable topics. Unlike other methods—where topic meaning must be inferred post hoc—in our

[23] For exposition, we display a selected subset of topics and include coarser, aggregated topic labels.

setting it is straightforward to link sentiment to topics and outlook, enabling both meaningful validation and clearer narrative interpretation.

4.3 Valuation Methods and Attention Allocation are Closely Linked

In a final step, we present initial evidence on the interrelation between valuation method choice and attention allocation, taking up the theoretical prediction that topic attention should vary across valuation methods depending on which variables are most relevant under the chosen framework.

Building on the cross-industry heterogeneity evidence from Panel D of Figure III, Appendix Figure VII shows that long-run variables such as corporate investment and innovation carry relatively greater weight in DCF-based reports than in those using multiples. More immediate variables—such as product- and customer-related topics—by contrast, are more commonly emphasized in multiples-based reports than in those using DCF.

Additionally, DCF forecasts tend to systematically incorporate different arguments—namely, a different time outlook. Specifically, in Columns 1 to 3 of Table III, DCF usage is associated with a reduced emphasis on backward-looking and near-term discussions, and greater focus on the distant future. These results are in line with the predictions of the framework that valuation method choice is interconnected with differential emphasis on more important aspects (i.e., forward-looking aspects for DCF methods explicitly projecting future cash flows). We defer the discussion of the final columns of Table III to Section 7.1.

Overall, a key result is the interdependence of the various components of analysts' mental models—with the choice of valuation methods both shaping and being shaped by analysts' attention allocation and lines of reasoning. We analyze the interaction between valuation methods and attention in further detail in Section 6.1.1 and Section 7.1.

5 From Mental Models to Quantitative Forecasts

In this section we study how the information we gathered about analysts' mental models is reflected into their quantitative forecasts. Valuation methods affect analysts' quantitative forecasts by definition (these are the formulas they use to compute their price targets).

Therefore, in this section we only focus on how analysts' attention allocation translates into their quantitative forecasts. We examine this both in the extensive margin (topic inclusion) and in the intensive margin (attention weights). Overall, we confirm that analysts' mental models are sparse, and we find that when analysts pay more attention to a topic category, their quantitative forecasts become more sensitive to variation in the underlying variable.

5.1 Extensive Margin: Mental Model Completeness

In Section 4, we provide evidence that analysts' mental models are sparse, in the sense that, on average, analysts only consider 18 out of 139 possible topics in their reports. To corroborate this interpretation, we conduct four complementary exercises.

First, we examine the longest decile of reports–those exceeding 4,700 words—to assess whether topic overlap increases when space constraints are minimal. Even in these instances, analysts covering the same firm-year diverge substantially in topic selection, as we discuss in Section 6.1.2. Second, we examine whether the observed sparsity reflects an artifact of the granularity of our 139-topic classification scheme. When we vary the granularity of our categories, sparsity remains prevalent even when we create categories with much lower topic similarity.[24]

Third, we ask whether there exist topics that are absent from an analyst report (or missed by the LLM extraction) yet nonetheless relevant for explaining analysts' quantitative forecasts. As we detail in the caption of Appendix Table E.5, we apply machine learning methods—elastic nets—to reports and find that, on average, topics omitted from a given report but included by other analysts covering the same firm-year are not selected by the algorithm as statistically significant predictors of an analyst's quantitative forecast (Column

[24] We assess how varying the granularity of topic definitions affects the extent of topical overlap. Specifically, we calculate the semantic similarity between topics using cosine similarity scores based on OpenAI-generated text embeddings of each topic labels. Appendix Table E.4 illustrates this procedure for the topic *antitrust*. We find that scores of 0.85 and above typically identify logically similar topic pairs. In contrast, scores around 0.80 often group conceptually distinct topics. We use these thresholds as benchmarks for evaluating the robustness of our findings. Panel E of Table I presents the distribution of Jaccard scores under thresholds of 0.85 and 0.80 for classifying a topic match. In both cases, we find that a substantial fraction of topics that do not perfectly overlap across analysts pairs. We also conduct a "kill test" to determine the level of topic coarsening required for at least 90% of analyst pairs to achieve a Jaccard score of 1. The resulting threshold is implausibly low, requiring the aggregation of topics with similarity scores as low as 0.76. This is akin to grouping *antitrust* with *commodity prices* and *housing demand*.

1 of Appendix Table E.5). To ensure that topics discussed by analysts are relevant, we perform the elastic net procedure on the topics that are included. On average, topics that are discussed in reports appear to be relevant in explaining the quantitative forecast (Column (1) of Panel B in Appendix Table E.5).[25]

Finally, we explore whether the sparsity of analysts' mental models is optimal or not. To do so, we analyze whether topics that are omitted by analysts can help explain their forecast errors. Intuitively, if forecast errors can be explained by omitted topics, but analysts' own valuations cannot, this would suggest that there are some topics that are excluded from analysts' mental models. To perform this exercise, we again apply an elastic net estimation. On average, the elastic net procedure does select some of the omitted topics as statistically relevant predictors of forecast errors (Column 2 of Appendix Table E.5). This suggests that the sparse nature of analysts' mental model may not be entirely optimal, and may instead be driven by limited cognitive resources.

5.2 Intensive Margin: Topic Attention and Sensitivity of Quantitative Forecasts

Next, we study how changing attention on the intensive margin influences analysts' quantitative forecasts. In particular, we are interested in understanding whether paying more attention to a given variable makes an analyst's forecast more sensitive to variation in that variable. Specifically, given the conceptual framework in Section 2, we have that:[26]

$$\mathbb{E}_t^i[p_{ft}] = \sum_{k=1}^{K} m_k(s_{kft} - \mu_{kft})a_{kft}^i + m_k\mu_{kft} \tag{19}$$

[25] Analysts' reputational or career concerns may also influence the content of their reports. They may include topics perceived as in-demand to attract attention—even if those topics are not central to their valuation logic—or omit topics that could be viewed as questionable to avoid reputational risk. While such strategic behavior is plausible in some cases, our analysis finds little evidence that such behavior is systematic in our setting.

[26] To see how we derived this expression notice that:

$$\mathbb{E}_t^i[p_{ft}] = \sum_{k=1}^{K} \mathbb{E}_t^i[p_{ft}] = m_k\mathbb{E}_t^i[x_{kft}] = m_k\left(\mu_{kft} + \frac{\tau_{skt}^i}{\tau_{skt}^i + \tau_{0kt}}(s_{kft} - \mu_{kft})\right) \tag{18}$$

Defining $a_{kft}^i \equiv \tau_{skt}^i/(\tau_{skt}^i + \tau_{0kt})$ and rearranging yields the desired expression. As a reminder, $(s_{kft} - \mu_{kft})$ is the news analysts receive about variable x_{kft}, and μ_{kft} is their prior belief about that variable.

where $a^i_{kft} \equiv \tau^i_{skt}/(\tau^i_{skt}+\tau_{0kt})$ is increasing in how much attention analysts pay to feature k. We are interested in measuring whether the sensitivity of price targets to an underlying variable varies with attention:

$$\left| \partial \left(\frac{\partial \mathbb{E}^i_t[p_{ft}]}{\partial (s_{kft} - \mu_{kft})} \right) \Big/ \partial a_{ikft} \right| = |m_k| > 0 \tag{20}$$

The challenge in estimating this quantity comes from the fact that while we observe a proxy for attention (a^i_{kft}), but do not always observe changes in the underlying variables associated with our categories ($s_{kft} - \mu_{kft}$). Therefore, to study this question we perform two exercises. First, we exploit coarser categories for which we do observed changes in the underlying. Second, we exploit an empirical strategy that uses information on the direction of signals to make progress in cases where we do not observe variation in the underlying variables.

5.2.1 Attention and Sensitivity to Top-Line Items

In this section we focus on variation in attention to top-line items. While coarser than the finer categories we have access to, variation in sales are not only observable, but are also an empirically dominant category in reasoning. To measure how analysts' sensitivity to top-line items varies with attention, we run the following regression, where β_3 captures our parameter of interest in (20):

$$\mathbb{E}^i_t[p_{ft}] \;=\; \beta_{0f} \;+\; \beta_1\, s_{ft} \;+\; \beta_2\, g_{ift} \;+\; \beta_3\, s_{ft}\, g_{ift} + \epsilon_{ift} \tag{21}$$

Table IV shows that, on average, analysts allocate 32% of their discussion to sales (Panel B of Table I), and variation in this share significantly predicts belief sensitivity. Greater attention to the sales channel is associated with analysts' price targets becoming more responsive to changes in sales[27]—that is, the relationship between firm sales and price target forecasts becomes increasingly positive as sales focus increases. This association remains

[27] We compute change in sales as the year-over-year sales ratio, such that firm-years with sales declines remain in the analysis when taking the natural logarithm. We winsorize sales growth at the 0.5th and 99.5th percentiles to account for extreme outliers. The results in Table IV are robust to using the non-winsorized variable.

positive (though somewhat attenuated) as we introduce progressively granular fixed effects: from no fixed effects in Column 1, to firm and year fixed effects in Column 2, and to firm-year fixed effects in Column 3.[28] The latter specification enables comparisons across analysts covering the same firm in the same year, thereby absorbing any potential confounds that would homogeneously influence analyst expectations for that firm at that time. In terms of magnitudes, at the mean sales growth, an interquartile increase in attention to sales is associated with a 7–36% increase in price target's sensitivity to sales, corresponding to 6–29% of the variable's interquartile range.

Overall, these results reveal a direct link between topic attention and quantitative forecasts.

5.2.2 Attention and Shadow Prices

In order to measure the interaction term of interest using the regressions described in (19), we need to be able to observe changes in the underlying variable $(s_{kft} - \mu_{kf})$. For most topics in our analysis, such changes are not observable (e.g., we do not have quantitative and observable values for changes in the legal environment, regulation, customer loyalty and retention). In what follows we perform a modified analysis that still allows us to understand how the sensitivity of price targets with respect to the underlying features varies with attention, even when we cannot observe movements in the underlying variable itself.

To do so, we exploit the fact that we observe the direction of signals via our sentiment measure, and estimate λ_k^+ and λ_k^- from the following expression:[29]

$$\mathbb{E}_t^i[p_{ft}] = \sum_{k=1}^{K} \lambda_k^+ a_{ikft}^+ + \lambda_k^- a_{ikft}^- + \epsilon_{kft} \tag{22}$$

where $a_{ikft}^+ = \mathbf{1}(s_t - \mu_t \geq 0) \times a_{ikft}$ and $a_{jkft}^- = \mathbf{1}(s_t - \mu_t < 0) \times a_{ikft}$. If λ_k^+ is greater than zero, an increase in attention to category k increases the price target more in response to a positive signal. Similarly, if λ_k^- is less than zero, an increase in attention to category k

[28] In Column 3, sales growth is not fully absorbed by the year fixed effects (representing calendar-year fixed effects) because, for maximum precision, we use the most recent sales figure available to each analyst. Since not all fiscal years correspond to calendar years, this yields within-calendar-year variation of the sales variable.

[29] It is worth pointing out that we only estimate coefficients at the category level, even though the unobserved signal varies at the category-firm-year level. This may cause issues of endogeneity as attention is likely correlated with how important the variable is.

decreases the price target more in response to a negative signal. Therefore, finding that $\lambda_k^+ > 0$ and $\lambda_k^- < 0$ would imply that analysts' responsiveness to news in a given category is amplified when they pay more attention to it.

Figure E.13 shows the resulting estimated coefficients for each category.[30] There are three points to notice. First, most λ_k^+ estimates are above zero, and most λ_k^- estimates are below zero, indicating that when analysts pay more attention to a category, their price targets do become more sensitive to variation in that category's underlying signals.

Second, the strongest and most statistically significant relationships appear in categories that analysts devote the most attention to. For instance, by interpreting results through equation (22), we see that product market competition receives a large share of attention (Figure VI), and a one-standard-deviation increase in attention to this category leads to a sizable increase in price targets (for both positive and negative signals). By contrast, legal issues see relatively little attention, and changes in that attention have no statistically significant impact on price target.

If we were able to directly estimate m_k instead of λ_k^+ and λ_k^-, this would suggest that analysts do indeed pay more attention to more important variables, consistent with the results from the theoretical framework we introduced in Section 2. However, one should be mindful when interpreting the absolute size of these coefficients. Differences in λ_k^+ and λ_k^- can reflect not only the true relevance of a category for valuation (m_k) but also the size of variation of a typical signal ($s_{kft} - \mu_{kf}$). A category prone to smaller news changes will naturally have a lower coefficient, even if it is quite important.

Overall, these results suggest that analysts' forecasts become more responsive to signals in a given category as their attention to that category increases. This effect is most pronounced in categories that receive a high share of analysts' attention, which could be because those categories exhibit larger or more frequent signals, or because they are genuinely more important to firm valuation.

[30] The regression we run takes pairwise differences between analysts i and j, and we further normalize the LHS to ensure stationarity and ease of comparison across firms. When $\lambda_{kft}^i = \lambda_{kft}^j$, this regression provides better identification.

5.3 Sentiment and Quantitative Price Target Levels

Finally, to see how the information in the text data of equity analysts' reports translates into their quantitative forecasts, Table II also shows that our measure of sentiment positively predicts analysts' subjective price target forecasts. In the raw data, a report that has fully-positive sentiment is associated with a 47% increase in the normalized price target relative to a fully-neutral report. This positive association remains strong when including firm fixed effects, thereby leveraging within-firm variation.

6 Drivers of Disagreement

So far we have shown that the properties of mental models that we measure meaningfully shape analysts' price targets. In this section we study the key drivers of disagreement. Specifically, given our theoretical framework, disagreement between two analysts j and k takes the following form:

$$\mathbb{E}_t^i[p_{ft}] - \mathbb{E}^j[p_{ft}] = \sum_{k=1}^{K} \lambda_{kft}^i(a_{kft}^i - a_{kft}^j) + (\lambda_{kft}^i - \lambda_{kft}^j)a_{kft}^j + \delta_{kft}^{ij} \tag{23}$$

where $\lambda_{kft}^i \equiv m_i(s_{kft} - \mu_{kf})$ captures how valuable feature k is in analyst i's valuation method, scaled by the size of the signal, and $\delta_{kft}^{ij} \equiv (m_k^i - m_k^j)\mu_{kf}$ captures how differences in how analysts value a feature apply to the whole signal, and not just to its innovation. The expression in (23) then clearly shows that analysts can disagree either because they use different valuation weights $(\lambda_{kft}^i - \lambda_{kft}^j)$ and/or because they use different attention weights across variables $(a_{kft}^i - a_{kft}^j)$. We consider both sources of variation in turn.

We start this section by documenting the differences in valuation methods and attention across analysts, and show how these differences are associated with disagreement in the quantitative forecasts of pairs of analysts covering the same firm at the same point in time. After that, we explore the relative importance of differences in valuation weights and differences in attention weights for forecast disagreement. Finally, we take a first step toward understanding which categories contribute most to disagreement. Overall, this analysis allows us to make drivers of disagreement interpretable.

6.1 Mental Models and Disagreement

This section documents differences in mental models across analysts, and shows how these differences correlate with disagreement. We consider both components of an analysts' mental model in turn (valuation methods and attention allocation), before turning to their interaction.

6.1.1 Valuation Methods and Disagreement

Columns 1 and 2 of Table V provide evidence that pairs of analysts who cover the same firm at the same point in time provide much closer quantitative forecasts when they rely on the same valuation method. This relationship holds when we include firm-year fixed effects, thereby leveraging only variation across analyst pairs within the same firm and year. The economic magnitudes are sizable. Using the same valuation method is associated with a 4–7% reduction in quantitative disagreement relative to the mean.

The remaining columns of Table V show that one way in which valuation method alignment predicts lower forecast disagreement is through an increased topic alignment. This finding once again highlight that choice of valuation method and attention weights are tightly interlinked. Specifically, in Columns 3 and 4, using the same valuation method is associated with significantly more similar topic focus, as reflected in higher Jaccard topic similarity scores. The coefficient estimate (multiplied by 100 for ease of exposition) implies that, relative to the sample mean, topic similarity across analyst pairs increases by 6% when analysts use the same valuation method. Columns 5 and 6 show that this valuation-method-driven alignment in reasoning extends to the specific arguments made—capturing the joint congruence of topic, valuation channel, time outlook, and sentiment. The estimated economic magnitudes are similar.

6.1.2 Attention and Disagreement

Panel D of Table I provides basic descriptive statistics on differences in attention allocation across forecasters, based on Jaccard topic similarity scores. These scores measure the degree of topic overlap between pairs of analysts covering the same firm in the same year (cf. equation

(17) for the Jaccard score formula), and they lie between 0 and 1.[31] The mean Jaccard score shows that, on average, only 32% of the combined topics used by two forecasters overlap. Therefore, there is substantial variation in the topics analysts choose to include, even when evaluating the same firm at the same time. Importantly, these differences do not appear to stem from space constraints in equity reports.[32] When we compute Jaccard scores for analyst pairs where both reports fall in the top 10% of the length distribution (i.e., among the least constrained by space), we continue to observe a large gap in topic coverage, with analysts only sharing 40% of their covered topics on average.

We next examine whether heterogeneity in topic focus across forecasters predicts disagreement in their subjective quantitative beliefs. As before, we form pairs of analysts covering the same firm in the same year and compute their topic overlap using Jaccard similarity scores. Additionally, we construct a more granular measure of Jaccard argument similarity—where an argument, as previously defined, is a combination of a topic, valuation channel, time outlook, and sentiment.[33]

Table VI presents the results from regressing forecast disagreement on topic and argument similarity. In Column 1, which includes no fixed effects, both Jaccard topic similarity and argument similarity negatively predict forecast disagreement. An interquartile increase in topic similarity between two reports is associated with a decrease in forecast disagreement by 8% relative to the baseline disagreement (cf. Panel C of Table I). Similarly, an interquartile increase in argument similarity corresponds to a 11% reduction. Since both similarity measures are included jointly in the regression, the latter reflects the incremental effect of alignment in reasoning above and beyond topic agreement alone. The significant association between topic and argument similarity and forecast agreement persists when we add various fixed effects (firm-employer pair fixed effects in Column 2 and firm-year fixed effects in Column 3) and when we restrict the sample to report pairs where the word counts are within 10% of each other (Columns 4 and 5).

[31] Scores in the table are multiplied by 100 for ease of exposition.

[32] Even if they did, such variation would still reflect heterogeneity in analysts' judgments about topic relevance and prioritization.

[33] Appendix Table E.3 confirms that different components of pairwise analyst arguments are positively correlated. Specifically, analyst pairs with more similar topic-related narratives also tend to express more similar narrative sentiment. The same holds for time outlook: when arguments share the same time horizon, their associated sentiment is significantly more aligned on average.

Finally, while valuation methods and attention allocation are interlinked—and their alignment across forecaster pairs tends to co-vary—Columns 6 and 7 of Table VI show that meaningful variation in attention allocation remains even among analyst pairs using the same valuation method. Moreover, this residual variation in topic-level attention continues to predict differences in beliefs. This result holds even with firm-year fixed effects, leveraging only variation across forecasters covering the same firm at the same time using the same valuation method. This result points to the influence of additional factors beyond valuation method choice in driving heterogeneity in attention and beliefs. For example, Table VII and Appendix Table E.6 show how bottom-up attention may contribute to disagreement by influencing how available information is to different analysts.[34,35,36]

In sum, the results from Table VI show that differences in topic attention play a central role in shaping the degree of co-movement in quantitative beliefs across forecasters.

6.2 Exploring the Relative Importance of Drivers of Disagreement

We end this section by exploring the relative overall importance of the sources of disagreement: is variation in beliefs more strongly associated with differences in valuation methods and shadow prices, or with differences in attention to underlying topic categories? Of course, we have shown above that both valuation methods and attention variation are predictive of disagreement—and the framework further underscores that the two are closely intertwined and jointly reflect the underlying valuation parameters, such as signal volatility and the cost of acquiring information. With this in mind, the ultimate goal is to speak to whether, on average and unconditionally (i.e., averaging across firms and industries), forecast differences

[34] This could work through differential costs of acquiring information in our model.

[35] Our data includes the identities of individual forecasters as well as their geographic locations, inferred from country codes in phone numbers. We use this data to assess whether analysts who share more similar experiences and exposures exhibit more similar reasoning and quantitative beliefs as well. Table VII shows a strong mental model alignment effect associated with geographic proximity. Analyst pairs located in the same country exhibit significantly greater similarity in both narrative topics and arguments, with economic magnitudes comparable to those observed for shared valuation techniques in Table V. As in that case, closer alignment in arguments is again associated with substantially greater agreement in quantitative beliefs among analysts sharing the same location.

[36] In Appendix Table E.6, we provide evidence on one concrete channel that appears to matter: inference from local inflation experiences. Higher inflation in the analyst's country is positively associated with their topic attention to inflation, controlling for the level of inflation in the country of the forecasted firm. However, this positive association only emerges when local inflation is sufficiently salient—specifically, when it exceeds the commonly referenced 2% threshold.

arise more from differences in valuation weights or through differences in attention weights.

As a first step in this direction, we estimate equation (23) by exploiting the fact that we are able to observe the sign of news via our measure of sentiment. Panel A of Figure VIII plots the estimated shadow price levels for each category, while Panel B of Figure VIII plots the estimated differences in shadow prices across analyst pairs covering the same firm at the same point in time.[37]

Starting from Panel B, we see that for many categories, we cannot reject that the estimated differences in shadow prices are zero. There are exceptions—such as labor market and payout (for the positive estimates only), and crisis-related topics—but overall, the evidence suggests that, on average, analysts often apply similar shadow prices to most categories, regardless of whether the sentiment is positive or negative. Moreover, notice that while we are not able to directly identify m_k (since our shadow prices capture $\lambda^i_{kft} = m_k(s_{kft} - \mu_{ft})$), we are still able to rule out the possibility that these effects may have been driven by the lack of news in certain categories ($s_{kft} - \mu_{ft} = 0$). If the zero coefficients in Panel B were driven by the size of the shock ($s_{kft} - \mu_{kf} \approx 0$), we would expect the coefficients in Panel A to also be zero, as they are scaled by the same quantity. The fact that most coefficients in Panel A are instead different from zero suggests that it must be differences in weights ($m^i_k - m^j_k \approx 0$) that are driving these results. Therefore, this pattern of mostly insignificant differences points to a small role of shadow prices in contributing to disagreement.

Turning to Panel A, we notice that differences in attention weights do instead contribute significantly to changes in disagreement. Moreover, these estimates are informative of how much each individual category contributes to total price target differences across analysts through differences in attention weights. Broadly, attention differences to more firm-specific categories—such as payout, product market competition, product, and brand—appear to matter the most, whereas attention differences to more macro-oriented or exogenous categories—such as interest rates, housing, or government—appear to matter less.

Overall, the above evidence represents a first step toward quantifying the variation in

[37] If we were able to observe variation in the underlying variables, we would be able to separately identify the true weights that analysts put on different variables (m^i_k) from the size of the news associated with each category ($s_{kft} - \mu_{ft}$). While we are not able to draw this distinction as we only observe $\lambda^i_{kft} = m_k(s_{kft} - \mu_{ft})$, the estimates in Figure E.13 are still informative of how much differences in attention and differences in valuation methods contribute to *total* disagreement.

belief disagreement into economically interpretable components. We find that differences in attention weights are a bigger source of disagreement than differences in valuation weights, and that it is attention to firm-specific categories that drive most of that variation. This also speaks to a literature in behavioral economics that studies the difference between representational complexity and computational complexity (Ba et al., 2022). We find that there is more agreement in how analysts overcome computational complexity (which valuation method to use as an approximation to the true valuation function) than how they overcome representational complexity (what to attend to given their chosen valuation method).

7 Further Discussion and Conclusions

In this section we discuss additional analysis on forecast errors and on asset pricing implications, and conclude.

7.1 Drivers of Forecast Errors

Given our framework, we can write forecast errors as:

$$P_{ft} - \mathbb{E}_t^i[P_{ft}] = \sum_{k=1}^{K} \lambda_{kft}^{true}(a_{kft}^* - a_{kft}^i) + (\lambda_{kft}^{true} - \lambda_{kft}^i)a_{kft}^i + \delta_{kf}^{itrue} + \xi_{ft} \tag{24}$$

where a_{kft}^* is the optimal level of attention in a rational benchmark, $\lambda_{kft}^i \equiv v_k^i(s_{kft} - \mu_{kf})$, $\delta_{kf}^{itrue} \equiv (v_k^{true} - v_k^i)\mu_{kf}$, where $\xi_{ft} \equiv P_{ft} - \mathbb{E}_t^*[P_{ft}]$. Systematic forecast errors relative to the rational benchmark may then be driven either from the analyst's choice of valuation method, or from deviations in attention weights from the rational benchmark.

7.1.1 Valuation Methods and Forecast Errors

Columns 4 and 5 of Table III show that DCF usage is associated with significantly lower forecast errors. This result holds both without fixed effects and with firm-year fixed effects, the latter comparing forecast errors for the same firm in the same year across analysts using different valuation methods. The economic effect sizes are meaningful, with DCF usage being associated with a 6% lower forecast error relative to the mean (cf. Panel C of Table I).

7.1.2 Attention and Forecast Errors

As discussed in Section 5.1, the sparsity of analysts' mental models may be suboptimal, as the topics they omit from their reports help predict forecast errors. However, attention to more prominent topics also contributes to forecast errors, as Table E.7 further shows that a greater focus to sales related variables is associated with larger forecast errors.

Making progress on how attention allocation across individual variables drives forecast errors is more challenging due to the fact that we do not observe the optimal level of attention a^*, nor do we observe the optimal valuation weight λ_{kft}^{true}. If we were able to assume that $\lambda_{kft}^{i} = \lambda_{kft}^{true} = \lambda_{kft}$ (i.e., analysts use the true valuation model on average), then the expression in (24) would reduce to:

$$P_{ft} - \mathbb{E}_t^i[P_{ft}] = \sum_{k=1}^{K} \lambda_{kft}(a_{kft}^* - a_{kft}^i) + \xi_{ft} \tag{25}$$

We would then be able to estimate the following equation:

$$P_{ft} - \mathbb{E}_t^i[P_{ft}] = \sum_{k=1}^{K} \Big(\lambda_k^+ \times \mathbf{1}(s_{kft} - \mu_{kf} > 0)\Big)\, \beta_k^+ + \Big(\lambda_k^- \times \mathbf{1}(s_{kft} - \mu_{kf} < 0)\Big) \times \beta_k^- \tag{26}$$

where the sign of β_k^+ and β_k^- would be informative about the prevalence of over or underreaction to positive or negative news for a given feature. Interestingly, observing the sign of a piece of news through our measure of sentiment would allow us to measure the prevalence of over- and underreaction in price targets, even without observing analysts' forecasts revisions (which we do not have for price targets). Conditional on knowing that analysts updated their beliefs upwards, a positive forecast error would be indicative of underreaction (the analyst updated upwards but not enough), while a negative forecast error would be indicative of overreaction (the analyst updated upwards but too much). We leave this exploration for future iterations, and note that the signed information we are able to extract from text data can allow us to make progress on this front.

7.2 Asset Pricing Implications

Having shown that our measures of analysts' mental models are meaningfully reflected in analysts' quantitative forecasts, and that they help explaining both disagreement and forecast errors, in this final section we provide a brief further discussion linking mental models to key asset pricing patterns. To do so, we start by focusing on the more aggregate information we gathered, and leave the analyses of how valuation methods and attention to finer categories influence asset pricing patterns to future iterations.

In the time-series, we find evidence of significant co-movement between sentiment in analyst reasoning and Shiller's CAPE index (Panel F of Table V). In the cross-section, Figure IX shows how average sentiment across analysts varies across bins in various cross-sectional sorts: CAPM beta, book-to-market, momentum, size, subsequent-year return, Capex investment, and profitability. Broadly, these patterns are consistent with those of realized returns. However, our dataset allows us to go beyond sentiment. Figure X shows how average time outlook varies across the same set of cross-sectional characteristics. Interestingly, time outlook is substantially more forward-looking for low growth firms than for value firms. Additionally, with the exception of the very left tail, time outlook is increasingly forward-looking in firm investment. Finally Figure XI shows even more granular patterns by looking at the difference in attention focus across categories for the highest and lowest sorts of each factor. Growth stocks attract more attention around product strategy, brand, and competition than value stocks. Similarly, large stocks attract more attention to payout policy and FX, reflecting potentially larger cash positions, more maturity, and a more international focus than small stocks.

Broadly, these associations between mental model and asset prices point to promising directions for linking the structure of mental models to the formation of market expectations and to the study of return predictability and mispricing in an interpretable way.

7.3 Conclusion

This paper analyzes the text of 1.6 million equity reports to measure the mental models underlying equity analysts' quantitative forecasts. We define mental models as a combination of valuation methods and attention weights over different arguments—with each argument defined by a topic and its associated valuation channel, time outlook, and sentiment. We show that the choice of valuation methods and the allocation of attention across variables are tightly interlinked, shaping variation in analysts' quantitative forecasts and disagreement, while also correlating with key asset pricing patterns.

References

Adam, K., and S. Nagel. 2023. Expectations data in asset pricing. In *Handbook of economic expectations*, pp. 477–506. Elsevier.

Amromin, G., and S. A. Sharpe. 2014. From the horse's mouth: Economic conditions and investor expectations of risk and return. *Management Science* 60:845–866.

Andre, P., I. Haaland, C. Roth, M. Wiederholt, and J. Wohlfart. 2024a. Narratives about the Macroeconomy. Tech. rep., SAFE Working Paper.

Andre, P., C. Pizzinelli, C. Roth, and J. Wohlfart. 2022. Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies* 89:2958–2991.

Andre, P., P. Schirmer, and J. Wohlfart. 2024b. Mental Models of the Stock Market. Tech. rep., Working Paper.

Asquith, P., M. B. Mikhail, and A. S. Au. 2005. Information content of equity analyst reports. *Journal of financial economics* 75:245–282.

Ba, C., J. A. Bohren, and A. Imas. 2022. Over-and underreaction to information. *Available at SSRN 4274617* .

Bacchetta, P., E. Mertens, and E. Van Wincoop. 2009. Predictability in financial markets: What do survey expectations tell us? *Journal of International Money and Finance* 28:406–426.

Bailey, M., E. Dávila, T. Kuchler, and J. Stroebel. 2019. House price beliefs and mortgage leverage choice. *The Review of Economic Studies* 86:2403–2452.

Bastianello, F. 2024. Time-series and cross-section of risk premia expectations: Evidence from Financial Analysts. *Available at SSRN 4204968* .

Bauer, R., K. Gödker, P. Smeets, and F. Zimmermann. 2024. Mental Models in Financial Markets: How Do Experts Reason about the Pricing of Climate Risk? *European Corporate Governance Institute–Finance Working Paper* .

Ben-David, I., and A. Chinco. 2024. Expected EPS× Trailing P/E. Tech. rep., National Bureau of Economic Research.

Bianchi, F., S. C. Ludvigson, and S. Ma. 2024. What hundreds of economic news events say about belief overreaction in the stock market. Tech. rep., National Bureau of Economic Research.

Binetti, A., F. Nuzzi, and S. Stantcheva. 2024. People's Understanding of Inflation. Tech. rep., National Bureau of Economic Research.

Bordalo, P., J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer. 2023. How people use statistics. Tech. rep., National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer. 2020a. *Expectations of fundamentals and stock market puzzles*. National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, G. Lanzani, and A. Shleifer. 2024a. A Cognitive Theory of Reasoning and Choice .

Bordalo, P., N. Gennaioli, R. LaPorta, and A. Shleifer. 2019. Diagnostic expectations and stock returns. *Journal of Finance* 74:2839–2874.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer. 2020b. Overreaction in macroeconomic expectations. *American Economic Review* 110:2748–2782.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer. 2024b. Finance without exotic risk. *Working paper* .

Bradshaw, M. T. 2002. The use of target prices to justify sell-side analysts' stock recommendations. *Accounting Horizons* 16:27–41.

Bybee, J. L. 2023. The ghost in the machine: Generating beliefs with large language models. *Working Paper* .

Bybee, L., B. Kelly, A. Manela, and D. Xiu. 2024. Business news and business cycles. *The Journal of Finance* 79:3105–3147.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu. 2020. The structure of economic news. Tech. rep., National Bureau of Economic Research.

Case, K. E., R. J. Shiller, and A. Thompson. 2012. What have they been thinking? Home buyer behavior in hot and cold markets. Tech. rep., National Bureau of Economic Research.

Charles, C., and C. Kendall. 2024. Causal narratives. *Available at SSRN 4669371* .

Charles, C., and P. Sui. 2024. Marketwide Memory. *Working paper* .

Chen, H., B.-H. Hwang, and Z. Peng. 2025. Why do investors like short-leg securities? Evidence from a textual analysis of buy recommendations. Tech. rep.

Chen, Z. 2024. Memory and Beliefs in Financial Markets: A Machine Learning Approach. *Working Paper* .

Chinco, A., S. M. Hartzmark, and A. B. Sussman. 2022. A new test of risk factor relevance. *The Journal of Finance* 77:2183–2238.

Chopra, F., and I. Haaland. 2023. Conducting qualitative interviews with AI .

De Bondt, W. F., and R. H. Thaler. 1990. Do security analysts overreact? *The American economic review* pp. 52–57.

De La O, R., and S. Myers. 2021. Subjective Cash Flow and Discount Rate. *Journal of Finance* .

De Rosa, L. 2024. (Deep) Learning Analyst Memory. Tech. rep., Working Paper.

Decaire, P., and M. Guenzel. 2024. What Drives Very Long-Run Cash Flow Expectations? *Working Paper* .

Decaire, P. H., and J. Graham. 2024. Valuation Fundamentals. *Working paper* .

Decaire, P. H., D. Sosyura, and M. Wittry. 2024. Resolving Estimation Ambiguity. *Working paper* .

Demirakos, E. G., N. C. Strong, and M. Walker. 2004. What valuation models do analysts use? *Accounting horizons* 18:221–240.

Engelberg, J., A. Manela, W. Mullins, and L. Vulicevic. 2025. Entity Neutering. Working paper.

Erkilet, G., G. Janke, and R. Kasperzak. 2022. How valuation approach choice affects financial analysts' target price accuracy. *Journal of Business Economics* 92:741–779.

Fan, T., Y. Liang, and C. Peng. 2021. The inference-forecast gap in belief updating. *Unpubished Manuscript* .

Flynn, J. P., and K. Sastry. 2024. Attention cycles. Tech. rep., National Bureau of Economic Research.

Gabaix, X. 2014. A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics* 129:1661–1710.

Gabaix, X. 2019. Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, vol. 2, pp. 261–343. Elsevier.

Gabaix, X., R. S. Koijen, R. Richmond, and M. Yogo. 2023. Asset embeddings. *Available at SSRN 4507511* .

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein. 2018. *Channeled attention and stable errors*. Harvard Business School Boston.

Giglio, S., and B. Kelly. 2018. Excess volatility: Beyond discount rates. *The Quarterly Journal of Economics* 133:71–127.

Giglio, S., B. Kelly, and D. Xiu. 2022. Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics* 14:337–368.

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

Gleason, C. A., W. Bruce Johnson, and H. Li. 2013. Valuation model use and the price target performance of sell-side equity analysts. *Contemporary Accounting Research* 30:80–115.

Gormsen, N. J., and K. Huber. 2023. Corporate discount rates. Tech. rep., National Bureau of Economic Research.

Graham, J. R. 2022. Presidential address: Corporate finance and reality. *The Journal of Finance* 77:1975–2049.

Greenwood, R., and A. Shleifer. 2014. Expectations of returns and expected returns. *The Review of Financial Studies* 27:714–746.

Haaland, I. K., C. Roth, S. Stantcheva, and J. Wohlfart. 2024. Measuring what is top of mind. Tech. rep., National Bureau of Economic Research.

He, S., L. Lv, A. Manela, and J. Wu. 2025. Chronologically Consistent Large Language Models. Working paper.

Hong, H., J. C. Stein, and J. Yu. 2007. Simple forecasts and paradigm shifts. *The Journal of Finance* 62:1207–1242.

Imam, S., J. Chan, and S. Z. A. Shah. 2013. Equity valuation models and target price accuracy in Europe: Evidence from equity reports. *International Review of Financial Analysis* 28:9–19.

Kacperczyk, M. T., S. van Nieuwerburgh, and L. Veldkamp. 2016. A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica* 84:571–626.

Ke, B. 2025. Analysts' Belief Formation in Their Own Words. Tech. rep., Working Paper.

Kohlhas, A. N., and A. Walther. 2021. Asymmetric attention. *American Economic Review* 111:2879–2925.

Laudenbach, C., A. Weber, J. Wohlfart, and R. Weber. 2024. Beliefs about the Stock Market and Investment Choices: Evidence from a Field Experiment. *Review of Financial Studies* .

Lopez-Lira, A., and Y. Tang. 2024. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *Working paper* .

Malmendier, U., and S. Nagel. 2011. Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics* 126:373–416.

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

Sarkar, S. K. 2025. Economic Representations. *Working Paper* .

Sarkar, S. K., and K. Vafa. 2024. Lookahead Bias in Pretrained Language Models. Tech. rep., Harvard Business School, Working Paper.

Schwartzstein, J., and A. Sunderam. 2021. Using models to persuade. *American Economic Review* 111:276–323.

Sims, C. A. 2003. Implications of rational inattention. *Journal of monetary Economics* 50:665–690.

Stantcheva, S. 2024. Why do we dislike inflation? Tech. rep., National Bureau of Economic Research.

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira. 2023. Man versus machine learning: The term structure of earnings expectations and conditional biases. *The Review of Financial Studies* 36:2361–2396.

Veldkamp, L. L. 2011. *Information choice in macroeconomics and finance.* Princeton University Press.

Vissing-Jorgensen, A. 2003. Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions. *NBER Macroeconomics Annual* 18:139–194.

Woodford, M. 2001. Imperfect common knowledge and the effects of monetary policy.

Figures and Tables

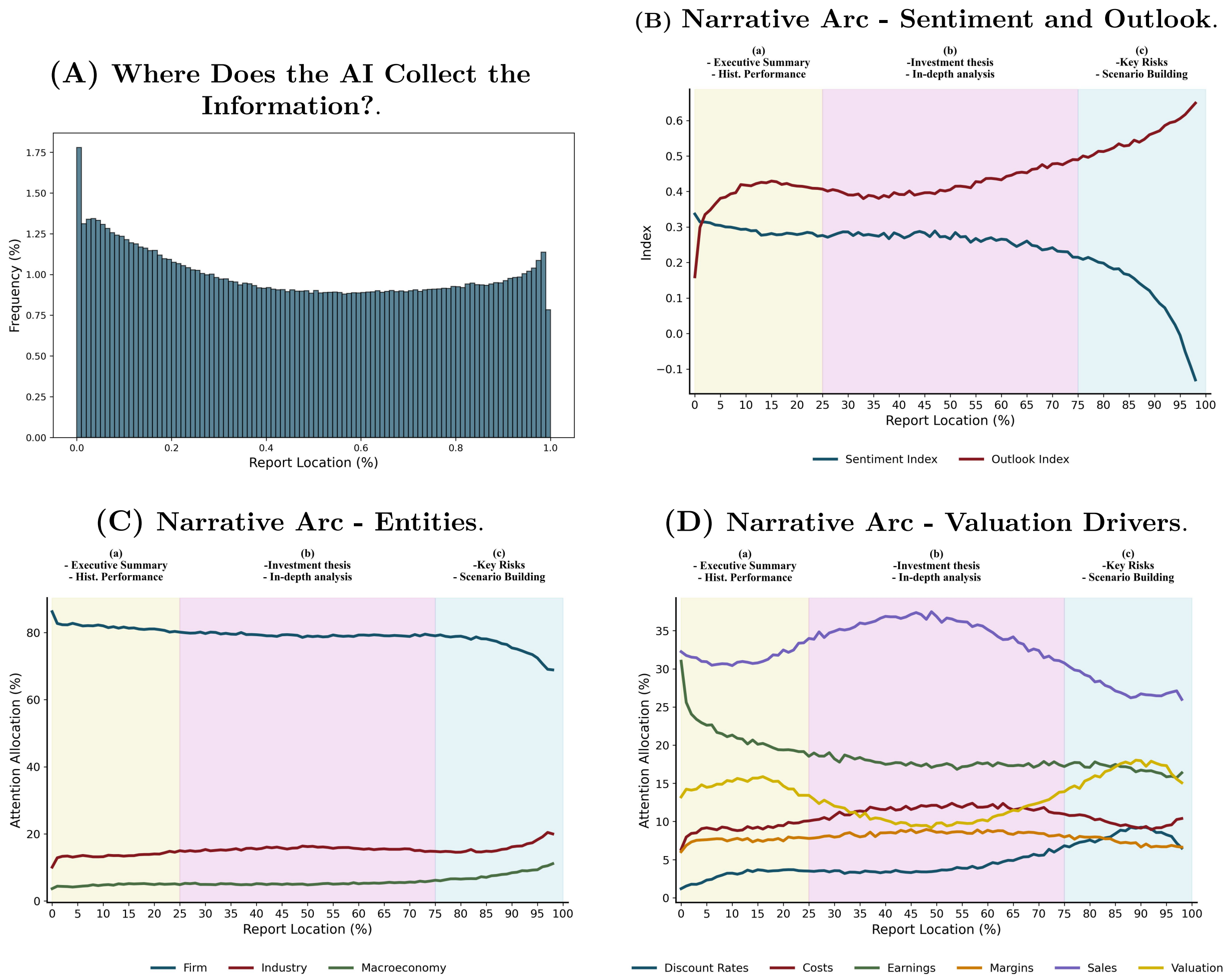


Figure I:
Report Narrative Arc

This figure shows where information is collected within the average report and the type of information collected at each stage. The x-axis for all four panels denotes the position of the report, in percentiles from start to finish. For example, a value of 1% corresponds to the topics discussed in the first 1% of a report. The sample period is 2000–2023 and includes all firms in our main sample. Panel A plots the share of topics extracted at each point in the report. Panel B plots average sentiment and time outlook over the course of reports. The *Sentiment Index* is calculated using values of -1, 0, and $+1$ for negative, neutral, and positive sentiment, respectively. The *Time Outlook Index* uses values of -1, 0, $+1$, and $+2$ for past, present, near-future, and distant-future outlook, respectively. Panel C plots the allocation of attention to entities (i.e., firm, industry, and the macroeconomy), while Panel D plots attention to valuation drivers (i.e., sales, costs, earnings/cash flows, margins, discount rates, and relative valuation statements). Shaded areas in the figures indicate the three commonly observed sections of report structure, providing intuitive reference points along the x-axis.

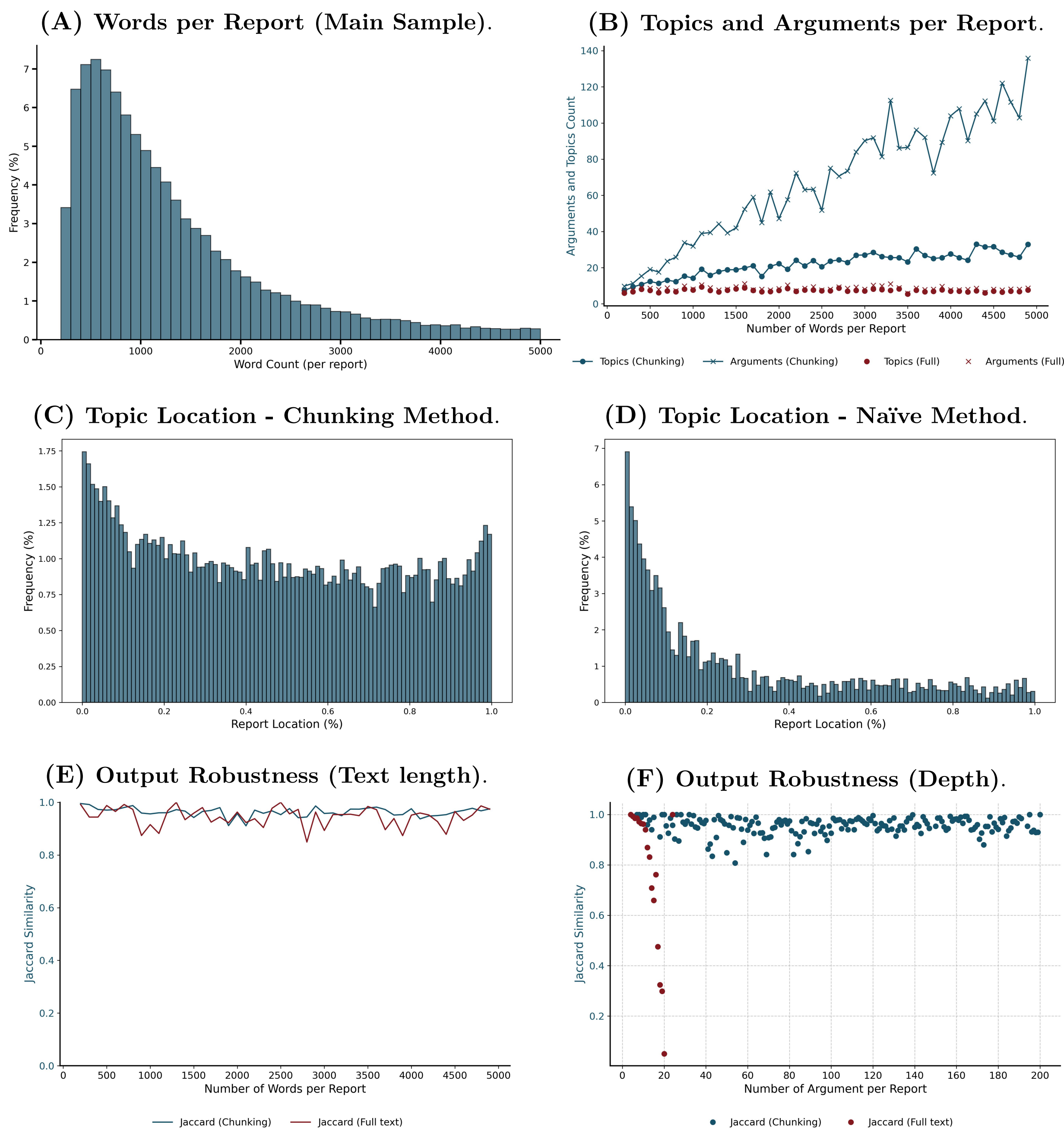
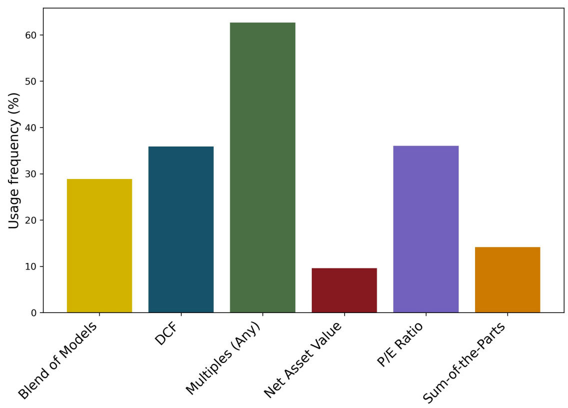


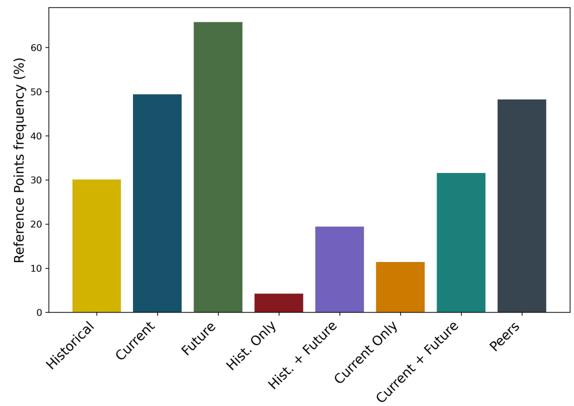
Figure II:
Diagnostic Tools

This figure presents results from diagnostic tests validating the reliability of our LLM-based extraction method. Panel A uses the main sample and plots the distribution of report word counts. Panels B through F use a subsample of 240 randomly selected reports (stratified by report length), each processed 10 times. Panel B shows the number of topics and arguments—defined as combinations of topic, valuation channel, time outlook, and sentiment—as a function of report length, comparing our multi-step extraction method (blue) with a naïve single-step method (red). Panels C and D plot the average position within the report at which topics are collected, for the multi-step and single-step approaches, respectively. Panel E reports the average Jaccard topic similarity across the 10 extraction attempts, by report length, comparing the two methods. Panel F does the same for Jaccard argument similarity.

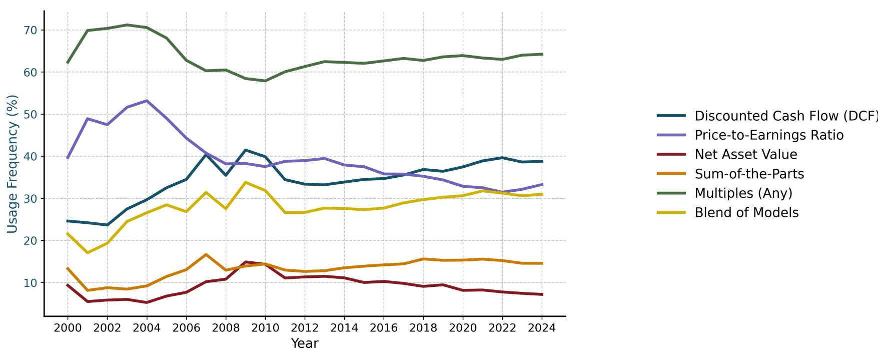
(A) Valuation Method Frequency per Report.



(B) Multiples Reference Point(s).



(C) By Year.



(D) By Industry (10 largest in sample).

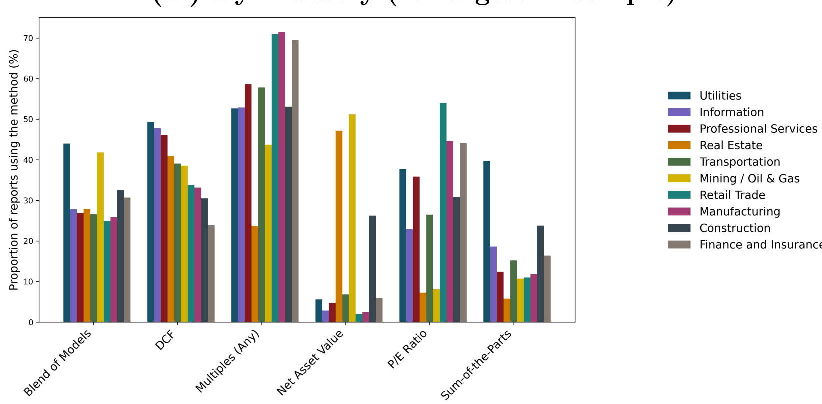


Figure III:
Valuation Methods

This table presents results on the use of valuation methods across the near-universe of equity reports in Refinitiv from 2000 to 2023 for which the LLM extraction identified the valuation method(s) used ($N = 1,243,873$ reports). Panel A reports the frequency of various valuation methods, including DCF, multiples, and blended approaches. Panel B presents statistics on reference points used in multiples-based valuations, distinguishing among historical, current, and future firm-based reference points (leftmost three columns) and peer-based reference points (rightmost column), along with several subcategories associated with firm-based reference points. Panels C and D show variation in the use of valuation methods over time and across industries.

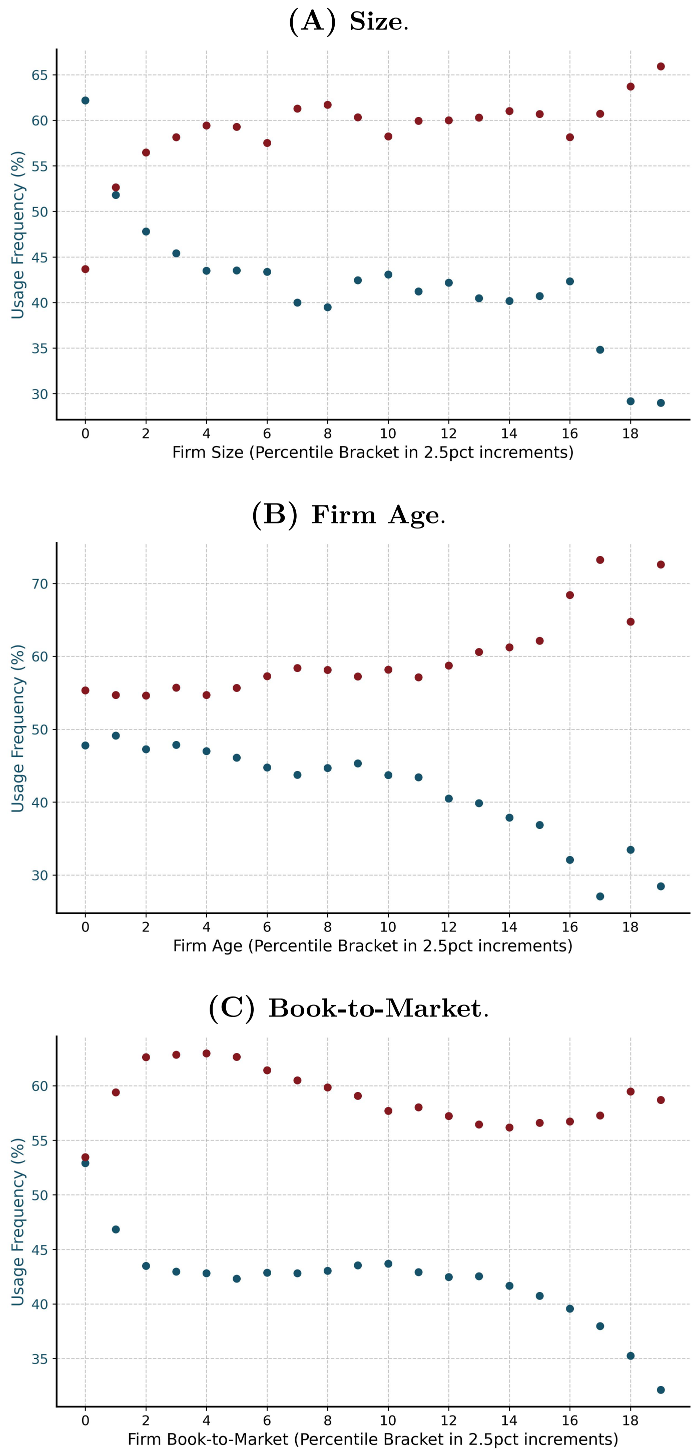


Figure IV:
Valuation Method Choice and the Cross-Section of Firms

This figure plots the propensity to use either the discounted cash flow (DCF) method or a multiples approach as a function of firm characteristics. Blue dots represents the DCF method, and red dots represent multiples-based methods. Panels A, B, and C show results by firm size (total assets), firm age (measured from the year of incorporation), and book-to-market ratio, respectively. For each characteristic, firms are sorted into 20 equal-sized bins (5%) within each year of the sample.

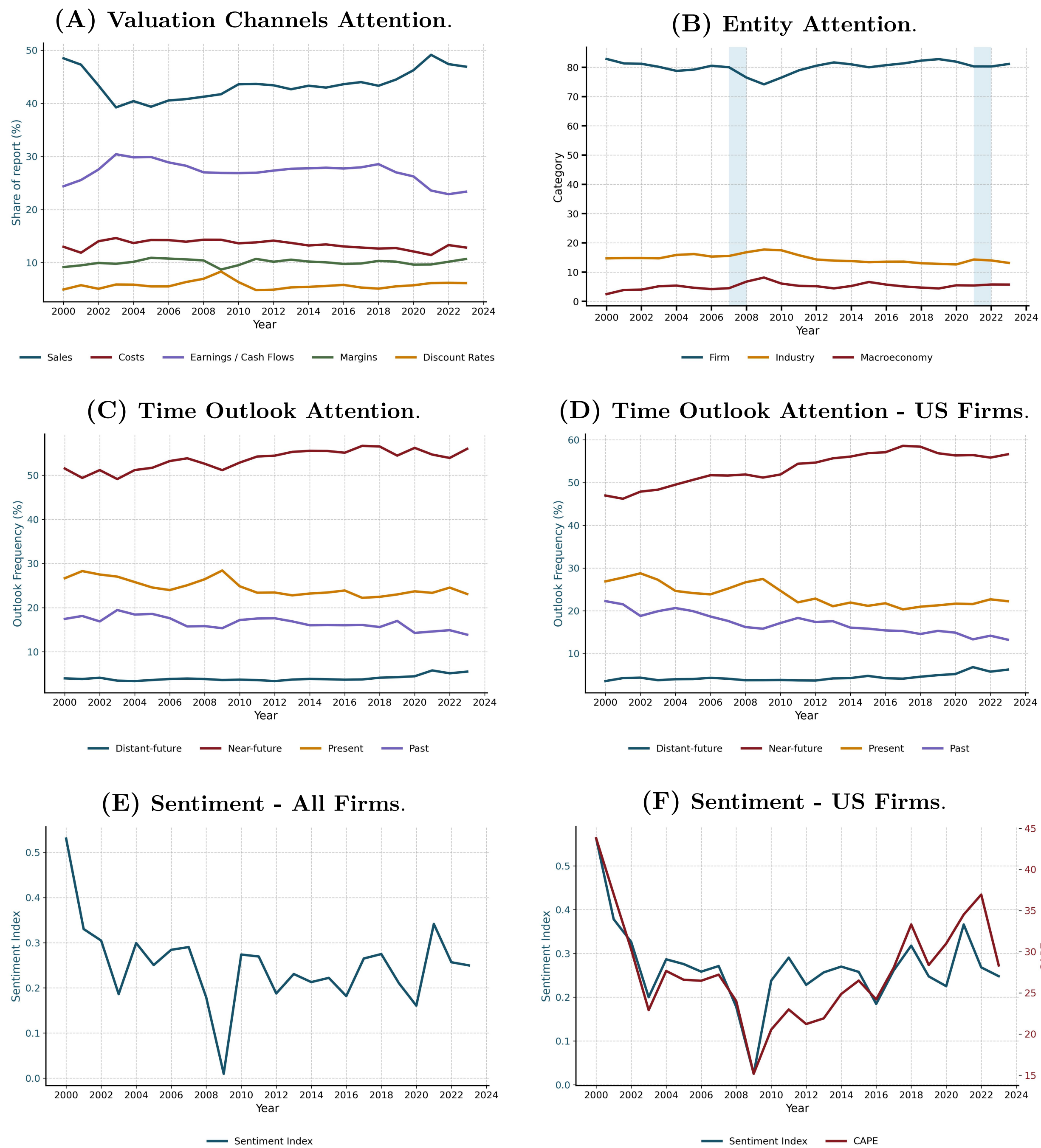


Figure V:
Valuation Channels, Entities, Outlook, and Sentiment

This figure plots time trends in mental model features across our sample, spanning 2000–2023. The x-axis denotes years in all panels. Panel A shows the share of statements in the average report that reference different valuation drivers—sales, costs, earnings, margins, and discount rates. Panel B reports the share of statements referencing the three entities: firm, industry, and macroeconomy. In Panels C and D, the y-axis indicates the percentage of statements in the average report that discuss topics related to the past, present, near future (<3 years), and distant future (>3 years), for all firms and U.S. firms only, respectively. Panels E and F display average sentiment across reports written in a given year, measured on a scale from -1 (negative) to $+1$ (positive), with 0 denoting neutral sentiment. In Panel F, the right-hand y-axis represents the scale for Shiller's CAPE index.

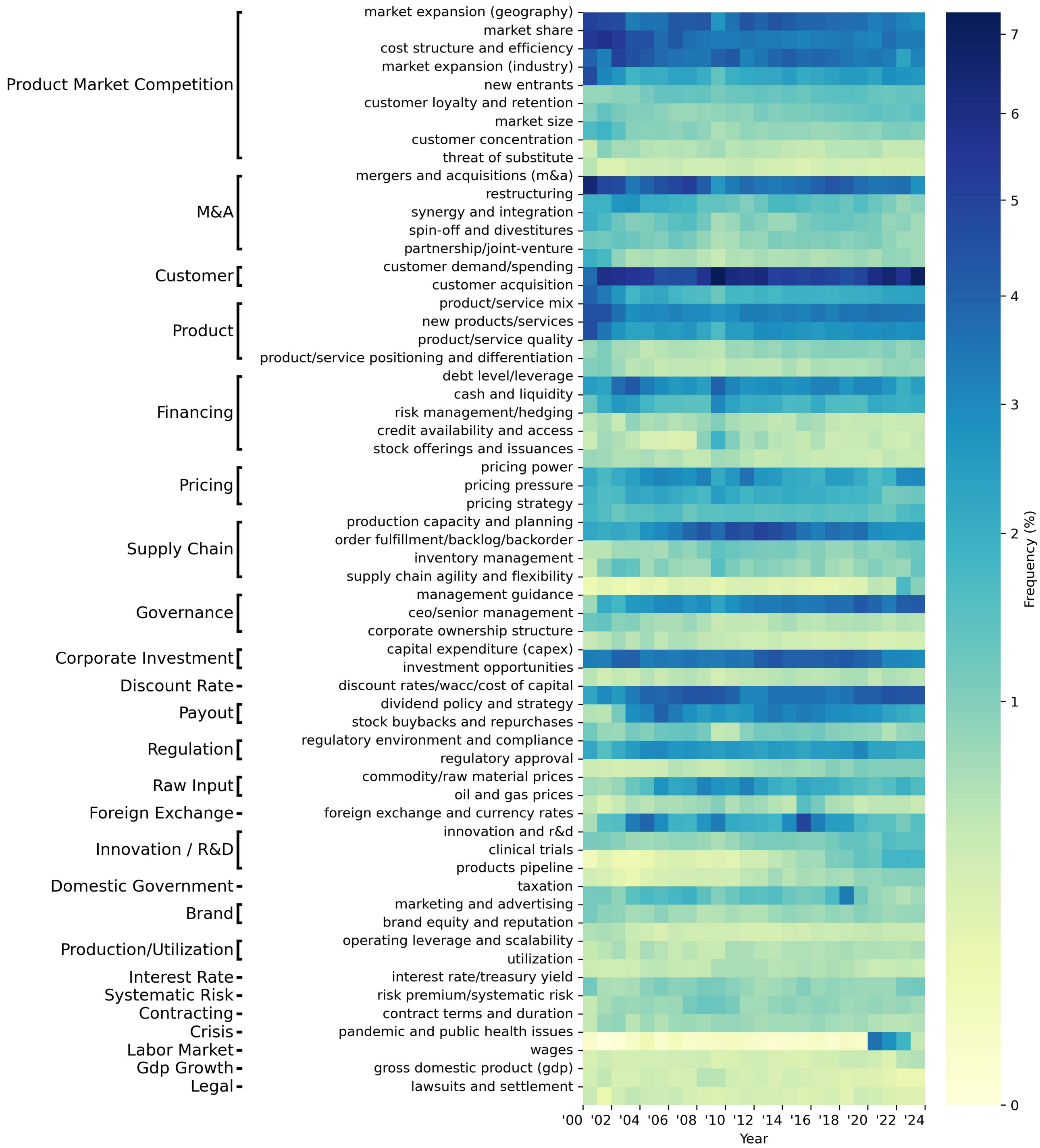


Figure VI:
Topic Attention Allocation Over Time

This figure plots attention allocation to the 60 most frequently discussed across all firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

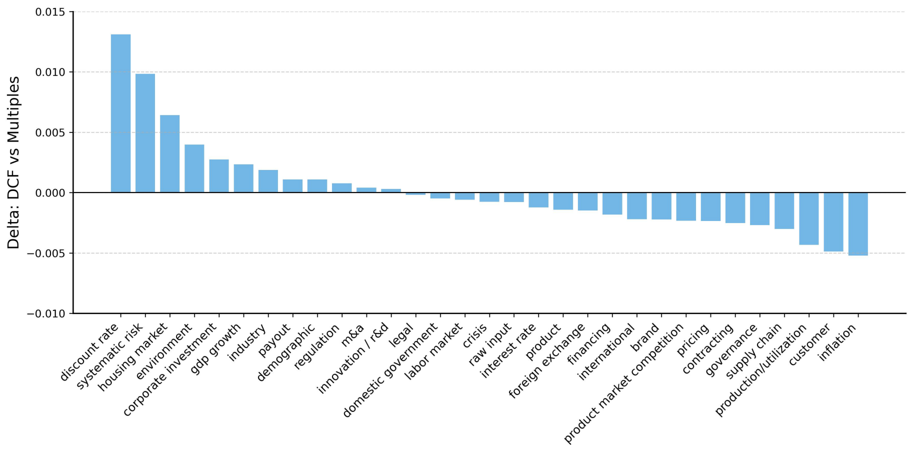
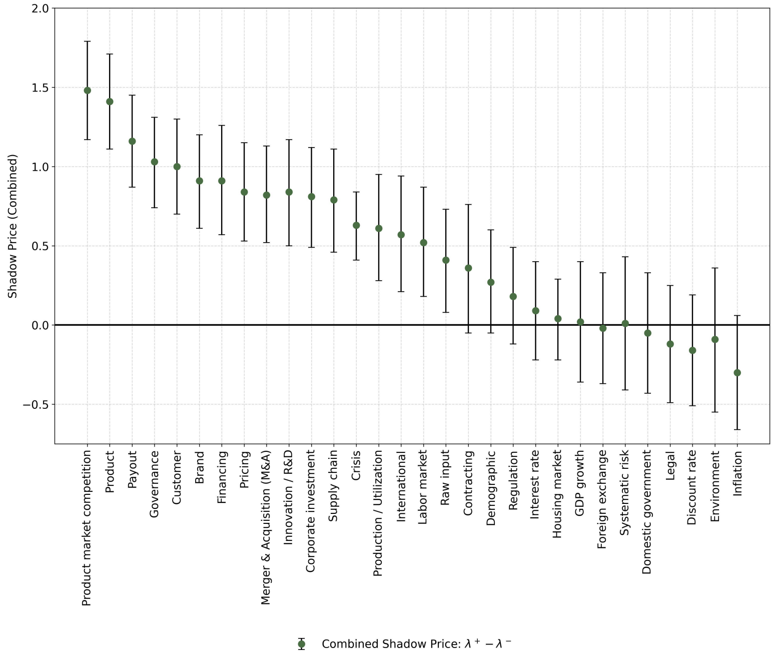


Figure VII:
Attention Allocation Delta: DCF Versus Multiples

This figure plots the average difference in attention weights allocated to topic categories between reports based on exclusively DCF models and those using exclusively multiples. A positive value indicates that DCF-driven valuations allocate more attention to that particular topic category, whereas a negative value denotes less attention than a multiple-driven approach.

(A) Shadow Price.



(B) Shadow Price Delta between Analyst Pairs.

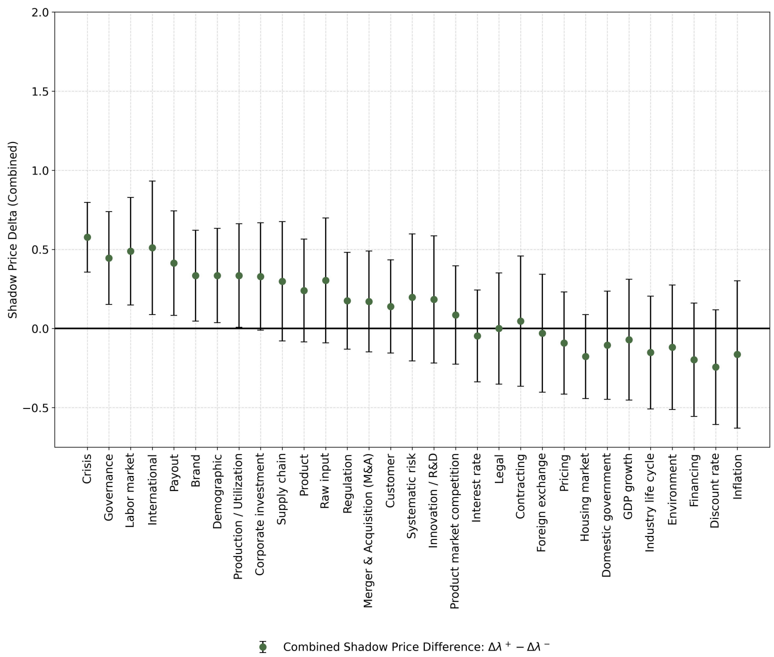


Figure VIII:
Shadow Price Estimates

This figure plots the shadow prices associated with each of the aggregate categories in our sample (see Appendix B). In Panel A, dots represent an average of the estimates of $\lambda^+_{k,B}$ and $\lambda^-_{k,B}$, respectively. In panel B, the dots represent an average of the estimates of $\lambda^+_{k,A} - \lambda^+_{k,B}$ and $\lambda^-_{k,A} - \lambda^-_{k,B}$, respectively. For both cases, the coefficients are jointly estimated from the regression $\frac{\mathbb{E}^i_t[p_{ft}]-\mathbb{E}^j_t[p_{ft}]}{\mathbb{E}^i_t[p_{ft}]+\mathbb{E}^j_t[p_{ft}]} = \sum_{k=1}^{K} \lambda^+_k \left(a_{ikft} - a_{jkft}\right)^+ + \lambda^-_k \left(a_{ikft} - a_{jkft}\right)^- + \left(\lambda^+_k - \lambda^+_k\right) a^+_{jkft} + \left(\lambda^-_k - \lambda^-_k\right) a^-_{jkft} + \epsilon_{ft}$ where $a^+_{k,i,j,t}$ ($a^-_{k,i,j,t}$) denotes the fraction of analyst i's report on firm j in year t allocated to topic k when discussed positively (negatively), and $\lambda^+_{k,i}$ ($\lambda^-_{k,i}$) are the associated shadow prices used by analyst i to price topic k scaled by the corresponding signal. The "undetermined" and "valuation" categories are excluded from the regression exercise. 90% confidence interval are measured using heteroskedastic consistent standard errors clustered at the firm level. $a^+_{k,i,j,t}$ and $a^-_{k,i,j,t}$ are scaled by their in-sample standard deviation to offer a direct measure of comparison across topics and topic sentiment direction.

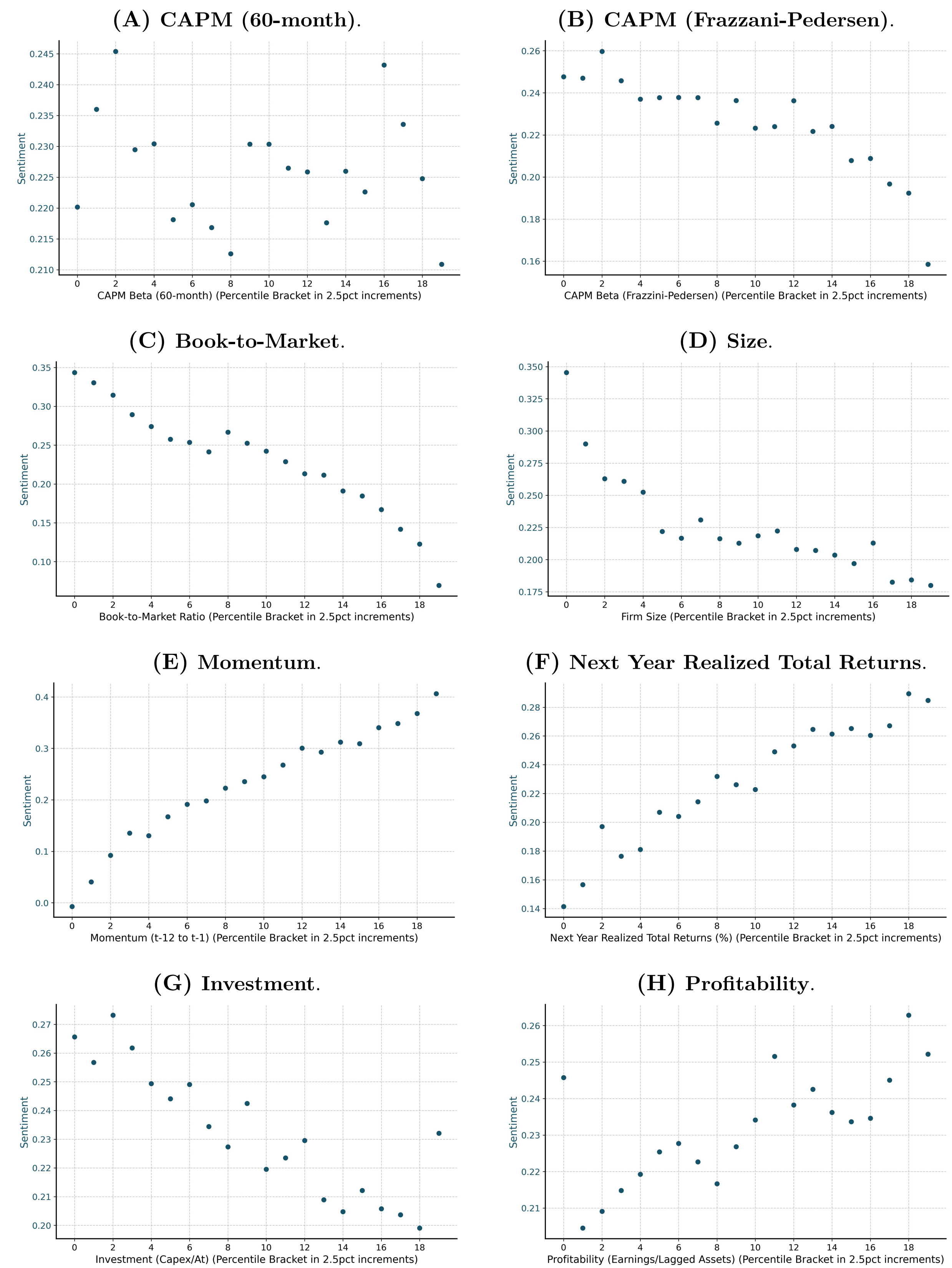


Figure IX:
Sentiment in the Cross-Section

This figure plots the cross-sectional patterns of sentiment for various key variables: CAPM beta, book-to-market, momentum, firm size, and birth cohort, across all the firms in our sample for which we have data on these characteristics. The x-axis corresponds to percentile brackets in 5% increments. For example, a value of 1 indicates the first percentile bracket, which includes observations where the x-variable falls within the 0 and 5 percentile range in each of their respective year. In all panels, the y-axis represents the average report sentiment within each percentile bracket. Report sentiment is calculated as the average sentiment of all topics discussed in the report, with values assigned as follows: -1 for negative sentiment, 0 for neutral sentiment, and 1 for positive sentiment. Larger values indicate more positive statements on average for the report.

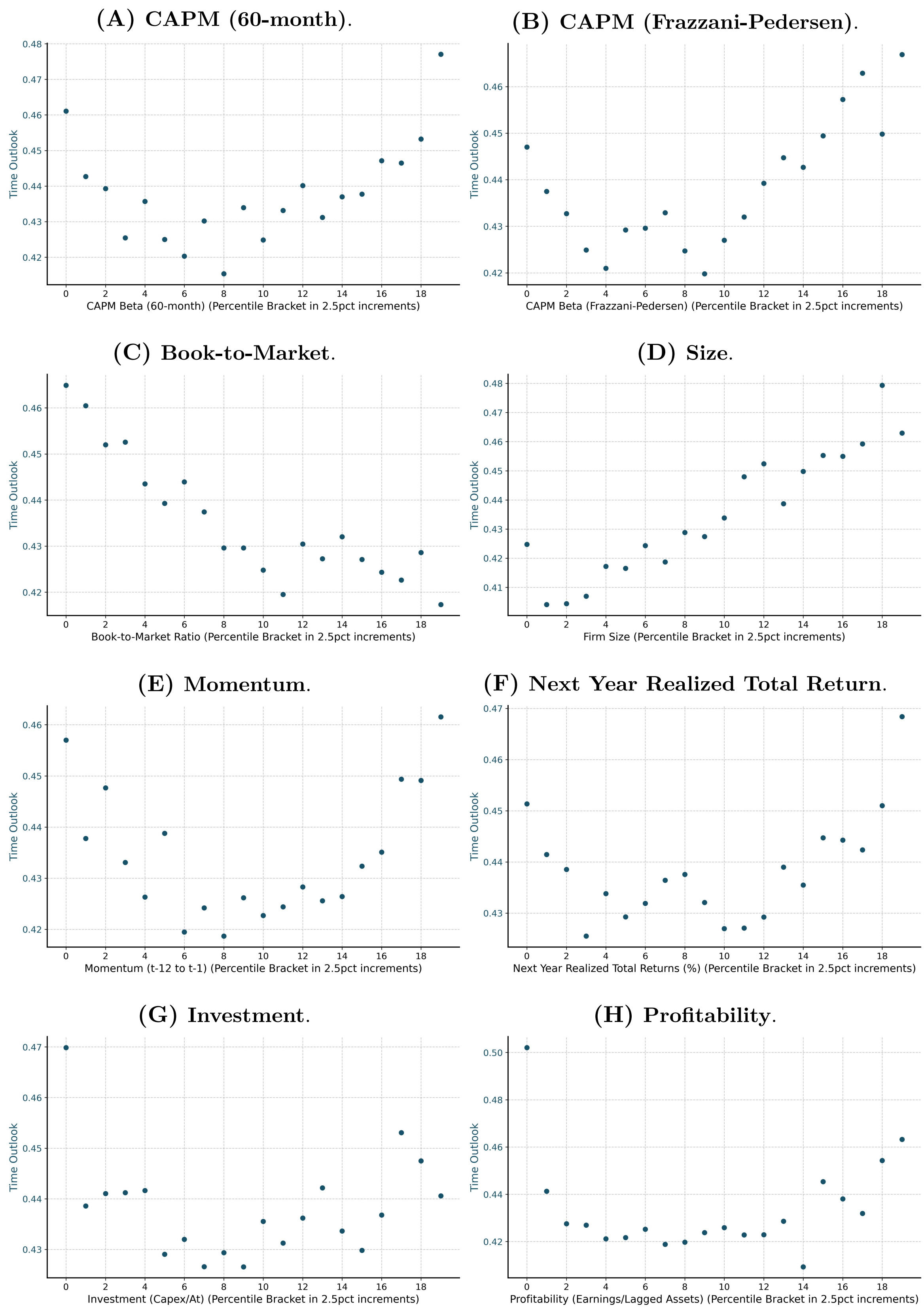


Figure X:
Time Outlook in the Cross-Section

This figure plots the cross-sectional patterns of time outlook for various key variables: CAPM beta, book-to-market, momentum, firm size, and birth cohort, across all the firms in our sample for which we have data on these characteristics. The x-axis corresponds to percentile brackets in 5% increments. For example, a value of 1 indicates the first percentile bracket, which includes observations where the x-variable falls within the 0 and 5 percentile range in each of their respective year. In all panels, the y-axis represents the average report time outlook in each percentile bracket. Report outlook is calculated as the average outlook of all topics discussed in the report, with values assigned as follows: -1 for past outlook, 0 for present outlook, 1 for near-future outlook, and 2 for distant-future outlook. Larger values indicate more forward-looking statements for the report.

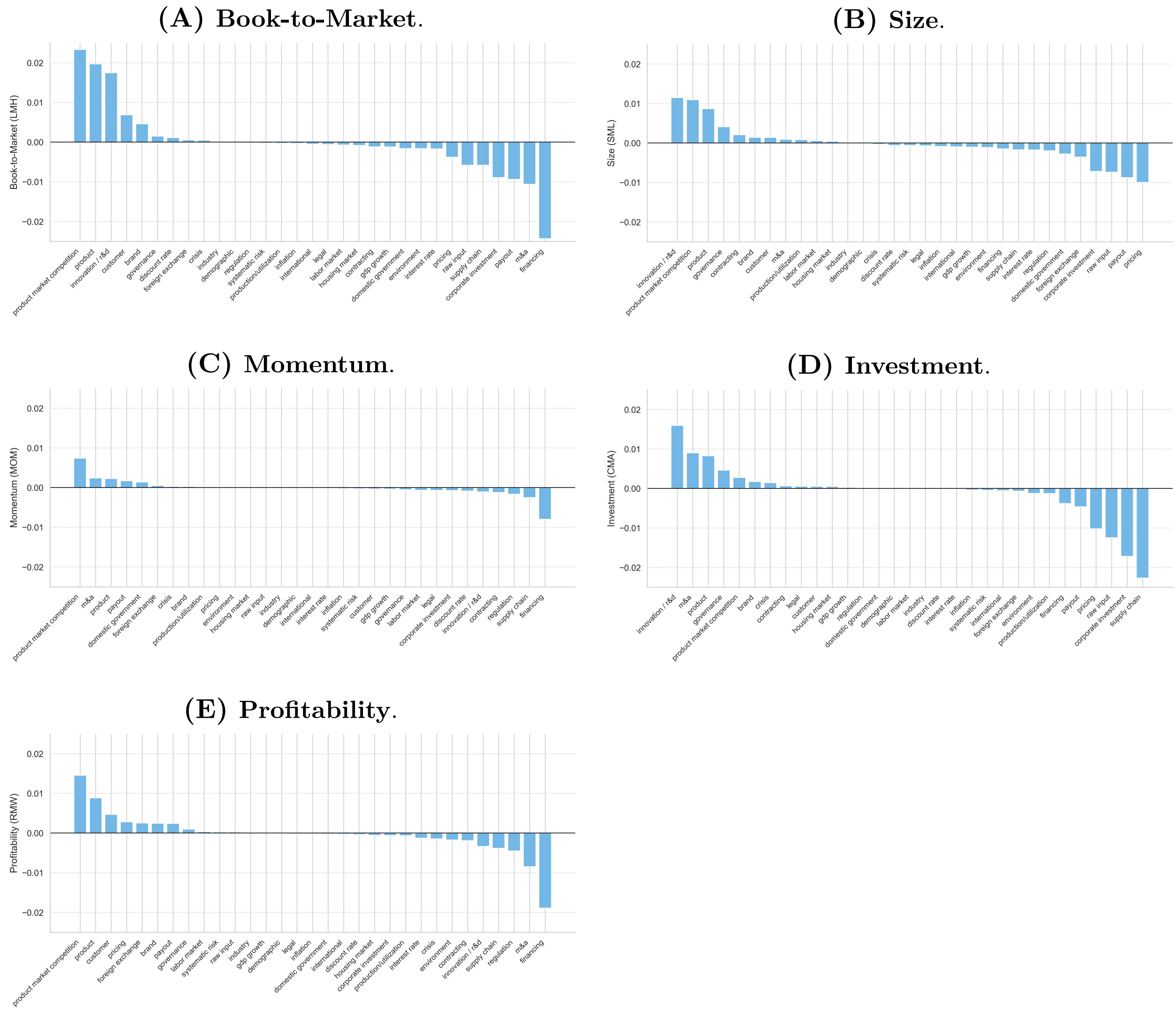


Figure XI:
Topic Focus in the Cross-Section

This figure plots the cross-sectional patterns of topic focus for various asset pricing sorts: book-to-market, momentum, firm size, profitability, and investment. The x-axis corresponds to categories. In all panels, the y-axis represents the average difference in how much a given category is discussed in the long-end and short-end of the corresponding asset pricing sort. Positive values indicate that a topic is discussed more on the long-end than the short-end.

Table I:
Descriptive Statistics on Equity Reports and Analysts' Mental Models

This table presents summary statistics based on our main sample of equity reports spanning 2000–2023. Panel A reports basic report characteristics. Panel B summarizes various features of extracted information, including topics, valuation channels, entities, sentiment, and time outlook. Panel C presents statistics on quantitative forecasts, forecast disagreement, and forecast errors. Panel D provides statistics on topic and sentiment similarity measures, while Panel E focuses on semantic similarity.

Panel A: Structure	Avg. Obs. per Category-Year	Avg. Obs per Category	Nb. Unique Category
Category:			
Equity report			111,270
Year		4636.25	24
Brokerage house	132.15	2587.67	43
Firm	2.19	9.84	11,308
Headquarters country	76.11	1163.61	90
Industry	231.18	5511.84	19

Panel B: Mental Model Features	Mean	25^{th} pct.	Median	75^{th} pct.	S. D.	Nb. Obs.
Report Text:						
Nb. Topic per report$_{i,j,t}$	17.85	13.00	17.00	22.00	6.46	111,270
Nb. Argument per report$_{i,j,t}$	42.79	24.00	36.00	55.00	25.34	111,270
Nb. Argument per topic (avg. report)$_{i,j,t}$	2.43	1.71	2.14	2.72	2.90	111,270
Nb. Word per report$_{i,j,t}$	1,304.37	605.00	1,008.00	1,689.00	963.29	111,270
Valuation Channels:						
Sales (%)$_{i,j,t}$	43.85	30.61	43.53	56.40	18.50	111,270
Costs (%)$_{i,j,t}$	13.25	6.74	11.39	17.75	8.65	111,270
Margins (%)$_{i,j,t}$	10.07	5.01	8.60	13.68	6.64	111,270
Earnings (%)$_{i,j,t}$	26.81	15.30	24.23	35.24	15.81	111,270
Discount rates (%)$_{i,j,t}$	5.94	2.45	4.28	7.68	5.26	111,270
Entities:						
Firm (%)$_{i,j,t}$	79.31	70.32	82.20	91.49	16.09	111,270
Industry (%)$_{i,j,t}$	15.09	6.88	12.39	20.65	11.07	111,270
Macroeconomy (%)$_{i,j,t}$	5.52	2.52	4.37	7.24	4.26	111,270
Sentiment:						
Sentiment$_{i,j,t}$	0.23	-0.05	0.28	0.56	0.42	111,270
Sentiment$^{past}_{i,j,t}$	0.27	-0.17	0.38	1.00	0.66	89,440
Sentiment$^{present}_{i,j,t}$	0.16	-0.25	0.20	0.62	0.59	105,481
Sentiment$^{near-future}_{i,j,t}$	0.24	-0.10	0.29	0.62	0.50	110,618
Sentiment$^{distant-future}_{i,j,t}$	0.44	0.00	0.67	1.00	0.65	57,353
Time Outlook:						
Outlook$_{i,j,t}$	0.47	0.25	0.51	0.72	0.35	111,270
Past (%)$_{i,j,t}$	17.31	6.99	14.19	24.74	12.93	111,270
Present (%)$_{i,j,t}$	22.32	13.55	20.57	29.18	12.27	111,270
Near-future (%)$_{i,j,t}$	55.01	42.62	55.24	67.36	17.83	111,270
Distant-future (%)$_{i,j,t}$	4.30	1.87	3.17	5.45	3.79	111,270
Panel C: Forecast Variables	Mean	25^{th} pct.	Median	75^{th} pct.	S. D.	Nb. Obs.
Forecast Error$_{i,j,t}$	0.16	0.05	0.10	0.18	0.19	67,223
$\ln(\frac{E_{i,j,t}[P_{i,t+1}]}{EPS_{j,t}})$	3.69	2.79	3.26	4.02	1.59	51,593
Disagreement$_{A,B,j,t}$	0.10	0.03	0.06	0.11	0.15	56,720
Panel D: Similarity Measures	Mean	25^{th} pct.	Median	75^{th} pct.	S. D.	Nb. Obs.
Topic Similarity (Analyst-Pairs):						
Jaccard$_{A,B,j,t}$	31.96	25.00	31.82	38.71	10.46	96,180
Jaccard$^{\text{90th pct < Report length}}_{A,B,j,t}$	39.89	33.33	39.47	45.65	9.72	1,291
Sentiment Agreement (Same Topic)						
Unconditional	0.48	0.00	0.00	1.00	0.50	3,672,802
Different argument & outlook & channel	0.44	0.00	0.00	1.00	0.50	811,717
Same argument & outlook & channel	0.73	0.00	1.00	1.00	0.45	48,458
Panel E: Semantic Similarity Measures	Mean	25^{th} pct.	Median	75^{th} pct.	S. D.	Nb. Obs.
Similarity (Analyst-Pairs) – 0.85 cutoff:						
Jaccard$_{A,B,j,t}$	43.64	34.78	44.00	52.63	13.02	96,180
Jaccard$^{\text{90th pct < Report length}}_{A,B,j,t}$	55.33	48.48	55.56	62.86	11.01	1291
Similarity (Analyst-Pairs) – 0.80 cutoff:						
Jaccard$_{A,B,j,t}$	85.52	80.77	87.50	92.31	9.71	96,180
Jaccard$^{\text{90th pct < Report length}}_{A,B,j,t}$	92.60	90.24	93.48	96.00	5.18	1,291

Table II:
Sentiment, Time Outlook, and 12-Month Price Target Forecasts

This table presents results examining the role of sentiment and time outlook in shaping analysts' 12-month price target forecasts. The analysis uses our main sample spanning 2000–2023. The dependent variable is the natural logarithm of analyst i's price target forecast for firm j in year t normalized by the firm's most recent earnings per share. $Sentiment_{i,j,t}$ represents the average sentiment of analyst i for firm j in year t. $Outlook_{i,j,t}$ denotes the average time outlook of analyst i for firm j in year t. Positive values for Sentiment and Outlook indicate positive sentiment and forward-looking outlook, respectively. $Sentiment_{i,j,t}^{Distant-future}$, $Sentiment_{i,j,t}^{Near-future}$, $Sentiment_{i,j,t}^{Present}$, and $Sentiment_{i,j,t}^{Past}$ are calculated using only the statements classified as referring to the distant future, near future, present, and past, respectively. Coefficients in Coefficients in Columns (4) to (7) are standardized to reflect one standard deviation changes in the independent variables, allowing for easier comparison across effects. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedasticity-consistent estimators clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	$\ln(\frac{E_{i,j,t}[P_{j,t+1}]}{EPS_{j,t}})$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(β_1) Sentiment$_{i,j,t}$	0.47***		0.19***				
	(0.03)		(0.01)				
(β_2) Outlook$_{i,j,t}$		0.31***	0.04***				
		(0.04)	(0.01)				
(β_3) Sentiment$_{i,j,t}^{Past}$				0.02***			
				(0.00)			
(β_4) Sentiment$_{i,j,t}^{Present}$					0.06***		
					(0.00)		
(β_5) Sentiment$_{i,j,t}^{Near-future}$						0.06***	
						(0.00)	
(β_6) Sentiment$_{i,j,t}^{Distant-future}$							0.02***
							(0.01)
Firm*Year FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	51,386	51,386	37,728	30,201	37,378	37,665	18,082
F Statistics	54.52	226.70	151.13	26.37	250.43	239.50	12.47
R^2	0.00	0.01	0.94	0.94	0.94	0.94	0.95

Table III:
Valuation Methods and Forecast Errors

This table presents results examining the role of valuation methods in explaining forecast errors. The analysis uses our main sample spanning 2000–2023. In Columns 1, 2, and 3, the dependent variables capture the extent to which analyst i's equity report for for firm j in year t relies on backward-looking, near-future, and distant-future arguments, respectively. In Columns 4 and 5, the dependent variable, *Forecast Error*, is defined as the absolute difference between forecaster i's 12-month price target and the realized price, divided by their sum, for firm j in year t. The key independent variable, *DCF Usage*$_{i,j,t}$ is an indicator variable that equals one if the analyst uses a DCF-based valuation method and zero otherwise. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedasticity-consistent estimators clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	Past$_{i,j,t}$	Near-future$_{i,j,t}$	Distant-future$_{i,j,t}$	Forecast Error$_{i,j,t}$	
	(1)	(2)	(3)	(4)	(5)
(β_1) DCF Usage$_{i,j,t}$	-0.01***	-0.03***	0.02***	-0.01**	-0.01***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Firm*Year FE	Yes	Yes	Yes	No	Yes
Observations	56,221	56,221	56,221	53,995	35,382
F Statistics	63.91	253.94	1,120.06	5.91	14.47
R^2	0.46	0.46	0.46	0.00	0.84

Table IV:
Topic Attention to Top-Line Items and 12-Month Price Target Forecasts

This table presents results examining how analysts' attention allocation to top-line items (as opposed to bottom-line items) affects the sensitivity of their 12-month price target forecasts to changes in sales. The dependent variable is the natural logarithm of analyst i's price target forecast for firm j in year t normalized by the firm's most recent earnings per share. The analysis uses our main sample spanning 2000–2023. Sales growth (measured as year-over-year sales ratio) is observed for firm j in month m of year t. Attention allocation to sales-related topics is measured for analyst i forecasting firm j in year t. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedastic-consistent estimators and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	$\ln(\frac{E_{i,j,t}[P_{j,t+1}]}{EPS_{j,t}})$		
	(1)	(2)	(3)
(β_1) ln(Sales growth)$_{j,m}$	0.17**	0.30***	0.40***
	(0.08)	(0.07)	(0.07)
(β_2) ln(Sales growth)$_{j,m}$*Attention Allocation$_{i,j,t}^{SalesChannel}$	1.18***	0.77***	0.23**
	(0.26)	(0.19)	(0.10)
(β_3) Attention Allocation$_{i,j,t}^{SalesChannel}$	-0.60*	-0.72***	-0.28**
	(0.32)	(0.21)	(0.12)
Firm FE	No	Yes	No
Year FE	No	Yes	No
Firm*Year FE	No	No	Yes
Observations	50,039	48,400	36,988
F Statistics	41.39	55.00	20.32
R^2	0.02	0.79	0.94
Within R^2	0.02	0.03	0.01

Table V:
Valuation Methods and Forecast Disagreement

This table presents results examining the role of valuation methods in explaining forecast (dis-)agreement. The analysis uses our main sample spanning 2000–2023. In Columns 1 and 2, the dependent variable, *Forecast* $Disagreement_{A,B,j,t}$, is defined as the absolute difference between forecaster A and B's 12-month price targets, divided by the sum of their respective price targets, for firm j in year t. These measures are calculated for analyst pairs evaluating the same firm in the same year. In Columns 3 and 4, the dependent variable, $Jaccard_{A,B,j,t}$, captures the number of overlapping topics between analyst A and B's reports divided by the total number of distinct topics mentioned in either report. In Columns 5 and 6, the dependent variable, *Same* $Argument_{A,B,j,t}$, is the share of overlapping topics for which both analysts assign the same time outlook, sentiment, and valuation channel, multiplied by 100. The key independent variable, *Same Valuation* $Method_{A,B,j,t}$, is an indicator variable equal to one if both analysts in the pair use the same valuation model (or the same combination of models) and zero otherwise. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedasticity-consistent estimators clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	Forecast $\text{Disagreement}_{A,B,j,t}$		$\text{Jaccard}_{A,B,j,t}$		Same $\text{Argument}_{A,B,j,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
(β_1) Same Valuation $\text{Method}_{A,B,j,t}$	-0.70***	-0.43**	1.81***	1.81***	0.29**	0.50***
	(0.19)	(0.19)	(0.13)	(0.14)	(0.12)	(0.13)
Firm*Year FE	No	Yes	No	Yes	No	Yes
Observations	36,039	28,291	36,039	28,291	36,031	28,285
F Statistics	14.04	4.88	195.60	161.57	6.19	14.27
R^2	0.00	0.53	0.01	0.40	0.00	0.36

Table VI:
Topic Alignment and Forecast (Dis-)Agreement

This table presents results examining the role of topic alignment in explaining forecast (dis-)agreement. The analysis uses our main sample spanning 2000–2023. The dependent variable, *Forecast Disagreement*, is defined as the absolute difference between forecaster A and B's 12-month price targets, divided by the sum of their respective price targets, for firm j in year t. The measure is calculated for analyst pairs evaluating the same firm in the same year. The key independent variables are $Jaccard_{A,B,j,t}$, which captures the number of overlapping topics between analyst A and B's reports divided by the total number of distinct topics mentioned in either report, and $Same\ Argument_{A,B,j,t}$, which is the share of overlapping topics for which both analysts assign the same time outlook, sentiment, and valuation channel, multiplied by 100. Columns 1 to 3 use the full sample. Columns 4 and 5 restrict the sample to analyst pairs whose reports differ in length by no more than 10 percent. Columns 6 and 7 restrict the sample to analyst pairs using the same valuation model (or the same combination of models), thus holding constant the methodology used to generate forecasts. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedasticity-consistent estimators clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	Forecast Disagreement$_{A,B,j,t}$						
Sample:	Main Sample			Both Reports are within a 10% Word Count Margin		Same valuation models	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(β_1) Jaccard$_{A,B,j,t}$	-0.06***	-0.05***	-0.03***	-0.05***	-0.08**	-0.04***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.03)
(β_2) Same Argument$_{A,B,j,t}$	-0.09***	-0.05***	-0.04***	-0.11***	-0.08*	-0.09***	-0.08**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.03)
Firm*Contributor$_A$*Contributor$_B$ FE	No	No	Yes	No	No	No	Yes
Firm*Year FE	No	Yes	Yes	No	Yes	No	Yes
Observations	56,660	47,720	19,103	5,124	1,747	16,655	2,642
F Statistics	83.15	43.49	15.27	26.24	5.92	40.78	6.39
R^2	0.01	0.51	0.81	0.01	0.61	0.00	0.81

Table VII:
Analyst Proximity and Forecast Disagreement

This table presents results examining the relation between analysts' country of location and topic focus alignment as well as forecast disagreement. The analysis uses our main sample spanning 2000–2023. In Columns 1 and 2, the dependent variable, $Jaccard_{A,B,j,t}$, captures the number of overlapping topics between analyst A and B's reports divided by the total number of distinct topics mentioned in either report. In Columns 3 and 4, the dependent variable, $Same\ Argument_{A,B,j,t}$, is the share of overlapping topics for which both analysts assign the same time outlook, sentiment, and valuation channel, multiplied by 100. In Columns 5 and 6, the dependent variable, $Forecast\ Disagreement_{A,B,j,t}$, is defined as the absolute difference between forecaster A and B's 12-month price targets, divided by the sum of their respective price targets, for firm j in year t. These measures are calculated for analyst pairs evaluating the same firm in the same year. The key independent variable, $1_{A,B,j,t}^{Analysts\ Located\ in\ the\ Same\ Country}$, is an indicator variable equal to one if both analysts are located in the same country and zero otherwise. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedastic-consistent estimators and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	$\text{Jaccard}_{A,B,j,t}$		$\text{Same Argument}_{A,B,j,t}$		$\text{Forecast Disagreement}_{A,B,j,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$(\beta_1)\ 1_{A,B,j,t}^{\text{Analysts Located in the Same Country}}$	1.05***	1.11***	0.67***	0.57***	-1.59***	-1.13***
	(0.17)	(0.20)	(0.16)	(0.19)	(0.34)	(0.34)
Firm*Year FE	No	Yes	No	Yes	No	Yes
Observations	40,595	32,995	40,574	32,976	40,595	32,995
F Statistics	39.75	30.83	18.82	9.11	22.04	11.02
R^2	0.00	0.39	0.00	0.36	0.00	0.53

Appendix A Proofs and Derivations

A.1 Derivation of Objective Function for Endogenous Valuation Method

The agent solves the following optimization problem:

$$\min_{m,\tau_{s1},\tau_{s2}} \quad \mathbb{E}\left[\left(p_t^{true} - \mathbb{E}_t[p_t^m]\right)^2\right] \qquad s.t. \qquad c_1\tau_{1s} + c_2\tau_{2s} = C \tag{27}$$

We can re-write the objective function as:

$$\mathbb{E}\Big[\Big(p_t^{true} - \mathbb{E}[p_t^m]\Big)^2\Big] = \mathbb{V}\Big[p^{true} - \mathbb{E}_t[p_t^m]\Big] + \mathbb{E}\Big[p_t^{true} - \mathbb{E}_t[p_t^m]\Big]^2 \tag{28}$$

$$= \mathbb{V}\left[\sum_{j=1,2}\Big[(v_j - m_j)\mathbb{E}_t[x_{jt}] + v_j\left(x_{jt} - \mathbb{E}_t[x_{jt}]\right)\Big]\right] + \mathbb{E}\Big[p_t^{true} - \mathbb{E}[p_t^m]\Big]^2 \tag{29}$$

$$= \sum_{j=1,2}(v_j - m_j)^2\mathbb{V}\left[\frac{\tau_{js}}{\tau_{js}+\tau_{0j}}\left(x_{jt} + u_{jt}\right)\right] + \sum_{j=1,2} v_j^2\left(\frac{1}{\tau_{sj}+\tau_{0j}}\right) + \mathbb{E}\Big[p_t^{true} - \mathbb{E}_t[p_t^m]\Big]^2 \tag{30}$$

$$= \sum_{j=1,2}(v_j - m_j)^2\left(\frac{\tau_{js}}{\tau_{js}+\tau_{0j}}\right)^2\left(\frac{1}{\tau_{0j}} + \frac{1}{\tau_{js}}\right) + \sum_{j=1,2} v_j^2\left(\frac{1}{\tau_{sj}+\tau_{0j}}\right) + \mathbb{E}\Big[p_t^{true} - \mathbb{E}_t[p_t^m]\Big]^2 \tag{31}$$

$$= \sum_{j=1,2}\frac{v_j^2}{\tau_{0j}+\tau_{js}} + \sum_{j=1,2}(v_j - m_j)^2\left(\frac{\tau_{js}}{\tau_{0j}(\tau_{0j}+\tau_{sj})}\right) + \left(\sum_{j=1,2}(v_j - m_j)\mu_j\right)^2 \tag{32}$$

$$= \sum_{j=1,2}\frac{v_j^2 - (v_j - m_j)^2}{\tau_{0j}+\tau_{js}} + \sum_{j=1,2}(v_j - m_j)^2\left(\frac{1}{\tau_{0j}}\right) + \left(\sum_{j=1,2}(v_j - m_j)\mu_j\right)^2 \tag{33}$$

where we used the fact that $\mathbb{V}[x_{jt} - \mathbb{E}_t[x_{jt}]] = (\tau_{js} + \tau_{0j})^{-1}$.[38] The final expression gives us the objective function in (13). Notice that when agents use the correct model, the second and third term cancel out and we are back to the objective function in (7).

[38] To derive this explicitly, we see that:

$$\mathbb{V}\left[x_{jt} - \mathbb{E}_t[x_{jt}]\right] = \mathbb{V}\left[x_{jt} - \frac{\tau_{sj}}{\tau_{sj}+\tau_{0j}}s_{jt} - \frac{\tau_{0j}}{\tau_{sj}+\tau_{0j}}\mu_j\right] = \mathbb{V}\left[x_{jt} - \frac{\tau_{sj}}{\tau_{sj}+\tau_{0j}}x_{jt} - \frac{\tau_{sj}}{\tau_{sj}+\tau_{0j}}u_{jt}\right] \tag{34}$$

$$= \mathbb{V}\left[\frac{\tau_{0j}}{\tau_{sj}+\tau_{0j}}x_{jt} - \frac{\tau_{sj}}{\tau_{sj}+\tau_{0j}}u_{jt}\right] = \left(\frac{\tau_{0j}}{\tau_{sj}+\tau_{0j}}\right)^2\frac{1}{\tau_{0j}} + \left(\frac{\tau_{sj}}{\tau_{sj}+\tau_{0j}}\right)^2\frac{1}{\tau_{sj}} = \frac{1}{\tau_{sj}+\tau_{0j}} \tag{35}$$

A.2 Proof of Proposition 1

We can either inspect the first order conditions in the text, or the optimal allocation of attention to variable x_k, which is given by the expression in (15):

$$\tau_{sk}^* = \frac{C + c_j\tau_{0j} - \sqrt{\frac{v_j^2}{v_k^2}}\sqrt{c_k c_j}\tau_{0k}}{c_k + \sqrt{\frac{v_j^2}{v_k^2}}\sqrt{c_k c_j}} \tag{36}$$

Ceteris paribus, this expression is clearly increasing in v_k (as the numerator is increasing in v_k and the denominator is decreasing in v_k), decreasing in c_k (as the numerator is decreasing in c_k and the denominator is increasing in c_k), and decreasing in τ_{0k} (as the numerator is decreasing in τ_{0k}). $\square$

A.3 Proof of Proposition 2

The marginal benefit evaluated at $\tau_{sk} = 0$ is given by: $v_k^2(\tau_{0k})^{-2}$. The marginal cost at $\tau_{sk} = 0$ λc_k. Since $\tau_{sk} \geq 0$, it follows that if the marginal cost is greater than the marginal benefit, then agents will optimally set $\tau_{sk} = 0$. Setting $\tau_{sk} = 0$ in the budget constraint then delivers $\tau_{sj} = \frac{C}{c_j}$.

Therefore for the optimal $\tau_{sk} = 0$, it must be that: $v_k^2(\tau_{0k})^{-2} < \frac{c_k v_j^2}{c_j\left(\tau_{02}+\frac{C}{c_j}\right)}$. Ceteris paribus, this condition is more likely satisfied when v_k is low, τ_{0k} is high, c_k is higher. This is also more likely satisfied when v_j is high, τ_{0j} is low,c_j is low, and when C is high. $\square$

A.4 Proof of Proposition 3

We are interested in proving that $\tau_{sk}^m > \tau_{sk}^* \iff \left(\frac{v_k - m_k}{v_k}\right)^2 < \left(\frac{v_j - m_j}{v_j}\right)^2$. Using the expressions in (11) and (15), we have:

$$\tau_{sk}^m > \tau_{sk}^* \iff \frac{C + c_j\tau_{0j} - \sqrt{\frac{v_j^2-(v_j-m_j)^2}{v_k^2-(v_k-m_k)^2}}\sqrt{c_k c_j}\tau_{0k}}{c_k + \sqrt{\frac{v_j^2-(v_j-m_j)^2}{v_k^2-(v_k-m_k)^2}}\sqrt{c_k c_j}} > \frac{C + c_j\tau_{0j} - \sqrt{\frac{v_j^2}{v_k^2}}\sqrt{c_k c_j}\tau_{0k}}{c_k + \sqrt{\frac{v_j^2}{v_k^2}}\sqrt{c_k c_j}} \tag{37}$$

This holds true if and only if:

$$\frac{v_j^2 - (v_j - m_j)^2}{v_k^2 - (v_k - m_k)^2} < \frac{v_j^2}{v_k^2} \iff \left(\frac{v_k - m_k}{v_k}\right)^2 < \left(\frac{v_j - m_j}{v_j}\right)^2 \tag{38}$$

$\square$

A.5 Proof of Proposition 4

Starting from a model with no wedges, we are interested in understanding when it might be optimal to introduce a wedge along a given dimension. Formally, we are interested in understanding conditions for which the following relationship holds:

$$MSE(w_1^2, 0) < MSE(0, 0) \tag{39}$$

where $w_1^2 = (v_1 - m_1)^2$. Using the definition of MSE in (16), the above condition holds if:

$$\frac{\left(\sqrt{c_1(v_1^2 - w_1^2)} + \sqrt{c_2 v_2^2}\right)^2}{C + c_1\tau_{01} + c_2\tau_{02}} + w_1^2\left(\frac{1}{\tau_{01}} + \mu_1^2\right) < \frac{\left(\sqrt{c_1 v_1^2} + \sqrt{c_2 v_2^2}\right)^2}{C + c_1\tau_{01} + c_2\tau_{02}} \tag{40}$$

$$w_1^2\left(\frac{1}{\tau_{01}} + \mu_1^2\right) < \frac{c_1 w_1^2 + 2v_2\sqrt{c_1 c_2}\left(\sqrt{v_1^2} - \sqrt{v_1^2 - w_1^2}\right)}{C + c_1\tau_{01} + c_2\tau_{02}} \tag{41}$$

For a given wedge w_1^2 and $v_1, v_2 \geq 0$, $v_1^2 \geq w_1^2$, this expression is easier to satisfy when the variable of interest is less valuable (lower v_1, μ_1), when the value of learning about the other variables is higher (lower τ_{02}, c_2, or higher v_2), or when the total budget on attention is lower (lower C). $\square$

Appendix B LLM Prompt to Extract Reasoning

Step 1: Identify key drivers of valuation exercise

Prompt input: Equity report.

Prompt output: Main mental model.

1. Main Task Analyze the provided equity report excerpt to identify all topics that clearly and directly explain the analyst's price target. For each topic:

- Categorize it into one of the following "category": "firm," "industry," or "macroeconomy."
- Categorize it into one of the following "channel": "earnings/cash flow: sales", "earnings/cash flow: costs", "earnings/cash flow: margins", "earnings/cash flow: earnings", "discount rate", "valuation", or "other".
- Identify its associated sentiment: "positive", "negative", "neutral", or "unclear."
- Identify the best keyword that captures the sentiment associated with the topic.
- Determine the time outlook: "past", "present", "near-future", "distant-future", or "unclear."
- Identify a 5 word snippet from the text to capture where the topic is first discussed.

2. Equity Report

```
<equityreport>
{EquityReport}
</equityreport>
```

3. Topic Identification

- Focus exclusively on topics directly tied to the justification of the price target.
- Each identified topic should be specific and clearly related to a single factor influencing the price target.
- Disregard general statements or any information not directly relevant to the price target justification.

4. Category Identification

- Assign each topic to the most appropriate category: "firm" (company-specific factors), "industry" (sector or industry trends), or "macroeconomy" (broader economic conditions).
- If a topic could reasonably belong to more than one category, select the category that best fits the context provided in the report.

5. Channel Identification Assign each topic to one of the following channels based on its context:

- "earnings/cash flow: sales": If related to sales or revenues.
- "earnings/cash flow: costs": If related to costs or operating expenses.
- "earnings/cash flow: margins": If related to profitability margins.
- "earnings/cash flow: earnings": If related to earnings, EBITDA, EBIT, or cash flows.
- "discount rate": If related to terms like "discount rate", "weighted average cost of capital/wacc", "equity beta", "idiosyncratic risk", "credit rating", "systematic risk", "equity risk premium (erp)", "market risk premium (mrp)", "interest rate", "risk-free rate (rf)", or "treasury yield".
- "valuation": If related to valuation terms such as "undervalued," "overvalued," or "fair-valued."
- "other": If none of the above apply.

6. Sentiment Identification Determine the sentiment associated with each topic, selecting one of the following:

- "positive": Indicates optimism or favorable conditions such as "beat expectations", "surpass expectations", "confidence", "bullish", "optimistic", ...
- "negative": Indicates pessimism or unfavorable conditions such as "uncertainty", "challenges", "caution", "concerns", "bearish", "underperform expectations", ...
- "neutral": Indicates neutrality or conditions in line with expectations such as "in line with", "meet consensus", ...

7. Keyword Identification

- Identify one or two keywords used by the analyst that best characterize the "Sentiment" variable.
- If no keywords can clearly characterize the sentiment, use "".

8. Time Outlook Identification Assign the time outlook based on the report's content. Remember that the report was written in the year {Year}. Possible values:

- "past": Events that have already occurred.
- "present": Events happening now.
- "near-future": Events expected within 1-3 years or short-term guidance.
- "distant-future": Events expected in more than 3 years or long-term guidance.
- "unclear": When the time frame is not clear.

9. Snippet Identification

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

10. Output Format For each topic, provide your answer in the following format:

```
{Topic}//{Category}//{Channel}//{Sentiment}//{Keyword}//{Timeoutlook}//{Snippet}
```

- `{Topic}` is the specific topic.
- `{Category}`: "firm," "industry," or "macroeconomy."
- `{Channel}`: "earnings/cash flow: sales", "earnings/cash flow: costs", "earnings/cash flow: margins", "earnings/cash flow: earnings", "discount rate", "valuation", or "other".
- `{Sentiment}`:"positive", "negative", "neutral", or "unclear."
- `{Keyword}`: the best keyword that captures the sentiment associated with the topic.
- `{Timeoutlook}`:"past", "present", "near-future", "distant-future", or "unclear."
- `{Snippet}`: 5 word snippet.

11. Example Here is an example of a text with the associated output to help you: "Investment summary: Inflation will raise the cost of inputs while also reducing demand. We have a negative outlook. Despite lagging behind its peers in the EV market, we remain optimistic about the firm's ability to increase its market presence over the next 5 years, as reflected in our hold recommendation. Pricing pressure in the company's main segment has compressed the firm's margins." In this example, the output should be:

```
inflation//macroeconomy//earnings/cash flow: costs//negative//negative
↪  outlook//near-future//Inflation will raise the cost
market share//firm//earnings/cash flow: sales//positive//remain
↪  optimistic//distant-future//Despite lagging behind its peers
pricing pressure//industry//earnings/cash flow:
↪  margins//negative//pressure//past//Pricing pressure in the company
```

This should be the only output you provide.

Step 2: Review the selection of valuation drivers and identify any missing relevant drivers

Prompt input: Equity report and previous analysis.

Prompt output: Omitted mental model.

1. Main Task Review the previous equity report analysis and identify all topics that clearly and directly explain the analyst's price target and that were omitted in the previous analysis, based on the provided equity report. For each omitted topic:

- Categorize it into one of the following "category": "firm," "industry," or "macroeconomy."
- Categorize it into one of the following "channel": "earnings/cash flow: sales", "earnings/cash flow: costs", "earnings/cash flow: margins", "earnings/cash flow: earnings", "discount rate", "valuation", or "other".
- Identify its associated sentiment: "positive", "negative", "neutral", or "unclear."
- Identify the best keyword that captures the sentiment associated with the topic.
- Determine the time outlook: "past", "present", "near-future", "distant-future", or "unclear."
- Identify a 5 word snippet from the text to capture where the topic is first discussed.

Importantly, do not output topics already included in the previous analysis.

2. Equity Report

```
<equityreport>
{EquityReport}
</equityreport>
```

3. Previous Analysis

```
<Previous Analysis>
{Initial Topics}
</Previous Analysis>
```

Previous analysis output structure is:

```
{Topic}//{Category}//{Channel}//{Sentiment}//{Keyword}//{Timeoutlook}//{Snippet}
```

Where

- `{Topic}` is a topic.
- `{Category}`: "firm," "industry," or "macroeconomy."
- `{Channel}`: "earnings/cash flow: sales", "earnings/cash flow: costs", "earnings/cash flow: margins", "earnings/cash flow: earnings", "discount rate", "valuation", or "other".
- `{Sentiment}`:"positive", "negative", "neutral", or "unclear."
- `{Keyword}`: the best keyword that captured the sentiment associated with the topic.
- `{Timeoutlook}`:"past", "present", "near-future", "distant-future", or "unclear."
- `{Snippet}`: 5 word snippet.

4. Omitted Topic Identification

- Focus exclusively on omitted topics directly tied to the justification of the price target.
- Each identified omitted topic should be specific and clearly related to a single factor influencing the price target.
- Disregard general statements or any information not directly relevant to the price target justification.

5. Category Identification

- Assign each omitted topic to the most appropriate category: "firm" (company-specific factors), "industry" (sector or industry trends), or "macroeconomy" (broader economic conditions).
- If an omitted topic could reasonably belong to more than one category, select the category that best fits the context provided in the report.

6. Channel Identification Assign each omitted topic to one of the following channels based on its context:

- "earnings/cash flow: sales": If related to sales or revenues.
- "earnings/cash flow: costs": If related to costs or operating expenses.
- "earnings/cash flow: margins": If related to profitability margins.
- "earnings/cash flow: earnings": If related to earnings, EBITDA, EBIT, or cash flows.
- "discount rate": If related to terms like "discount rate", "weighted average cost of capital/wacc", "equity beta", "idiosyncratic risk", "credit rating", "systematic risk", "equity risk premium (erp)", "market risk premium (mrp)", "interest rate", "risk-free rate (rf)", or "treasury yield".
- "valuation": If related to valuation terms such as "undervalued," "overvalued," or "fair-valued."
- "other": If none of the above apply.

7. Sentiment Identification Determine the sentiment associated with each omitted topic, selecting one of the following:

- "positive": Indicates optimism or favorable conditions such as "beat expectations", "surpass expectations", "confidence", "bullish", "optimistic", ...
- "negative": Indicates pessimism or unfavorable conditions such as "uncertainty", "challenges", "caution", "concerns", "bearish", "underperform expectations", ...
- "neutral": Indicates neutrality or conditions in line with expectations such as "in line with", "meet consensus", ...

8. Keyword Identification

- Identify one or two keywords used by the analyst that best characterize the "Sentiment" variable.
- If no keywords can clearly characterize the sentiment, use "".

9. Time Outlook Identification Assign the time outlook based on the report's content. Remember that the report was written in the year {Year}. Possible values:

- "past": Events that have already occurred.
- "present": Events happening now.
- "near-future": Events expected within 1-3 years or short-term guidance.
- "distant-future": Events expected in more than 3 years or long-term guidance.
- "unclear": When the time frame is not clear.

10. Snippet Identification

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

11. Output Format For each topic that is not already included in the previous analysis, provide your answer in the following format:

```
{Topic}//{Category}//{Channel}//{Sentiment}//{Keyword}//{Timeoutlook}//{Snippet}
```

- `{Topic}` is the specific omitted topic.
- `{Category}`: "firm," "industry," or "macroeconomy."
- `{Channel}`: "earnings/cash flow: sales", "earnings/cash flow: costs", "earnings/cash flow: margins", "earnings/cash flow: earnings", "discount rate", "valuation", or "other".
- `{Sentiment}`:"positive", "negative", "neutral", or "unclear."
- `{Keyword}`: the best keyword that captures the sentiment associated with the topic.
- `{Timeoutlook}`:"past", "present", "near-future", "distant-future", or "unclear."
- `{Snippet}`: 5 word snippet.

12. Example Here is an example of a text with the associated output to help you: "Investment summary: Inflation will raise the cost of inputs while also reducing demand. We have a negative outlook. Despite lagging behind its peers in the EV market, we remain optimistic about the firm's ability to increase its market presence over the next 5 years, as reflected in our hold recommendation. Pricing pressure in the company's main segment has compressed the firm's margins." In this example, the output should be:

```
inflation//macroeconomy//earnings/cash flow: costs//negative//negative
↪  outlook//near-future//Inflation will raise the cost
market share//firm//earnings/cash flow: sales//positive//remain
↪  optimistic//distant-future//Despite lagging behind its peers
pricing pressure//industry//earnings/cash flow:
↪  margins//negative//pressure//past//Pricing pressure in the company
```

This should be the only output you provide.

Step 3: Generate Standardized Labels for the Topics

Prompt input: Equity report, combined previous analysis, standardized topic lists.

Prompt output: standardized labels.

1. Main Task Identify the standardized labels that best match **each** topic from the list below.

- It is crucial to label every topic from the Topic List.

3. Previous Analysis

```
<topic_list>
{Topic List}
</topic_list>
```

Previous analysis output structure is:

```
{Topic}//{Sentiment}//{Timeoutlook}//{Snippet}
```

Where

- `{Topic}` is a topic.
- `{Sentiment}`:"positive", "negative", "neutral", or "unclear."
- `{Timeoutlook}`:"past", "present", "near-future", "distant-future", or "unclear."
- `{Snippet}`: is a 5 word snippet to help you identify where in the report this topic is discussed.

3. Equity Report

```
<equityreport>
{EquityReport}
</equityreport>
```

4. Instructions 1

- Use the "Standardized Label List 1" to label each topic if there is a clear and direct match in the following list:

```
<Standardized Label List 1>
{Standardized Label 1}
</Standardized Label List 1>
```

To apply the standardized labels "undervalued," "overvalued," and "fair-valued," the topic must involve a comparison of the current valuation with either the firm's peers or a historical reference point.

- Undervalued: Use this label when the topic suggests the stock is priced lower than its peers or historical benchmarks. Look for expressions such as "inexpensive," "undervalued," "at a discount," "cheap," or "trades below."
- Overvalued: Apply this label when the topic indicates the stock is priced higher than its peers or historical norms. Key expressions include "expensive," "overvalued," "trades at a premium," or "trades above."
- Fair-Valued: Use this label when the topic implies the stock's valuation is in line with its peers or historical standards. Look for phrases like "compares to," "similar to" or "trades on par."

If none of the labels are a clear and direct match for the topic, try this list instead as a last resort:

```
<Standardized Label List 2>
{Standardized Label 2}
</Standardized Label List 2>
```

Important: It is crucial to only use the labels provided in those two lists.

5. Refinement

- Organize the final list logically.

6. Output Format For each topic, provide your answer in the following format:

```
{Topic}//{Standardized Label}
```

- `{Topic}` is the specific omitted topic.
- `{Standardized Label}`: is the label that best matches the topic.

This should be the only thing that you output!

12. Example

```
Increased sales in Asia//sales and revenues
New competitors//new entrants
US GDP growth//gross domestic product (gdp)
```

make sure that you labeled all the topics contained in topic list.

Appendix C LLM Prompt To Extract Valuation Methods

//=== 1. Identify Valuation Methods ===

From the text, determine whether any of the following valuation methods are mentioned:

- Discounted Cash Flow (DCF) Analysis
- Price-to-Earnings Ratio (P/E or PER)
- Price-to-Book Ratio (P/B)
- Price-to-Sales Ratio (P/S)
- Enterprise Value to EBITDA (EV/EBITDA)
- Price-to-Cash Flow (P/CF)
- Net Asset Value (NAV)
- Liquidation Value
- Replacement Cost Method
- Sum-of-the-Parts (SOTP)
- Economic Value Added (EVA)
- Revenue Multiples (SaaS & Tech Startups)
- Other Valuation Method (If so, please specify.)

For each method, return `1` if present and `0` if absent.

—

//=== 2. Identify Primary Valuation Methods Used for Price Target ===

Determine which valuation methods explicitly contribute to the price target.

- Note: if more than one valuation method is directly tied to the price target, list all that apply.
- Note: if a valuation method is only used for reference (e.g., "supports" or "sanity check"), do not include it here.
- Note: if a multiple is used in combination with growth rates or discount rates, classify it under both the relevant multiple and DCF.

—

//=== 3. Identify Multiples Used & Their Comparisons ===

If multiples are mentioned, determine their **comparison basis**:

- **"Current Stock Price Value":** if the multiples or price target is compared to the company's **current trading multiple or stock price**.
- **"Historical Market Values":** if the the multiple is compared to **historical trends or past company trading levels**.
- **"Future Outlook":** if the multiple is based on **future growth expectations, discount rates, or DCF valuation.**
- **"Peer Valuation":** if the multiple is compared to those from **peer companies**.
- **"Industry Comparison":** if the multiple is compared to **sector-wide/industry averages**.

- Note: when evaluating "Future Outlook" do **not** classify based on how the multiple is applied — only consider **why it was chosen**, and **whether its justification is forward-looking**.

—

//=== 4. Snippet ===

Provide a snippet from the equity analyst report (up to two sentences) to accompany each of the answers you provide to the questions above.

- If the snippet is unavailable, return `"null"`.

—

//=== Examples ===

Primary Valuation Method Examples

– **Note:** Only include the valuation used to **generate** the price target directly.

- **Example:** "Our DCF-based PT implies 3.2x/2.7x '23/'24E EV/sales" — the valuation method used to derive the price target is **DCF only**, and not EV/Sales as well.
- **Example:** "Our valuation is based on DCF and supported by ~16x our 2014 EPS." — the valuation method used to derive the price target is **DCF only**, and not P/E as well.
- **Example:** "Our $165 price target is derived from a blend of two valuation techniques, equally weighted: 1) relative P/E valuation, which yields a value of $177 per share, and 2) our discount cash flow (DCF) model, which yields a value of $153." — the valuation methods used are **both DCF and P/E**.

– **Note:** Mentions of NPV or inputs to DCF should be tagged as **DCF**.

- **Example:** "We added $15 per share of additional value to reflect the estimated net present value (NPV)" — the valuation method used to derive the price target is **DCF**.
- **Example:** "We apply a long-term (post-tax) weighted average cost of capital (WACC) of 9.7%" — the valuation method used to derive the price target is **DCF**.

Multiples Comparison Examples

– Multiples Comparison Examples

- **Example:** "Price objective $58 implies CY09E P/E and EV/EBITDA multiples of 16.3x and 9.0x, respectively, nearer historical highs of 16.9x and 8.4x, reflecting more optimism regarding its ROIC potential." — set `multiples_comparison` = {Historical Market Values, Future Outlook}.
- **Example:** "Our SOTP applies an EV/2024E Revenue valuation of 1.7x for Mobility, 3.0x revenue for Delivery (implies 0.6x bookings), and 1.0x revenue for Freight, which are slight premiums to peers given potential network effects." — set `multiples_comparison` = {Peer Valuation, Future Outlook}.
- **Example:** "Our $75 PO is based on 18.6x our 2012E EPS, which is in line with the company's average 1-yr forward multiple since 2003, and is backed by a DCF analysis assuming 10% WACC and 3% terminal growth rate." — set `multiple_comparison` = {Historical Market Values, Future Outlook}.
- **Example:** "As stated earlier, we believe that the stock is currently overvalued with an earnings multiple of 122x our C03 estimate and 50x our C04 estimate." — set `multiple_comparison` = {Current Stock Price Value}.
- **Example:** "Our 8x target multiple comes from its recent historical average." — set `multiple_comparison` = {Historical Market Values}.
- **Example:** "We apply a multiple of 24x to our 2021 EPS estimate of $7.38." — set `multiple_comparison` = {null}.

– **Note:** Whenever the report mentions that shares **trade** at a given value, tag this as `"Current Stock Price Value"`.

- **Example:** "Univar shares trade on FY 2022E EV/EBITDA of 8.8x." — set `multiple_comparison` = {Current Stock Price Value}.
- **Example:** "Looking to 2010 EBITDA, NBL trades 3.9x vs. 3.5x for the group." — set `multiple_comparison` = {Current Stock Price Value, Peer Valuation}.
- **Example:** "AMT's current valuation (16x 2016 AFFO/share) lags REIT peers' (~20x), despite peer-leading AFFO/share growth (+13% in 2016/2017) and a rapidly growing dividend (~2.3% yield, growing 20%+ yoy)." — set `multiple_comparison` = {Current Stock Price Value, Future Outlook, Peer Valuation}.

– **Note:** "Future Outlook" should only be assigned when the multiple is explicitly justified by future trends, expected growth rates, or discount rates. It should **not** be used if the text simply states that the multiple is applied to a future estimate without providing a justification of how the multiple was generated.

- **Example (Future Outlook should be included):** "Given strong tailwinds from regulatory shifts and a projected 5-year EBITDA CAGR of 12%, we believe a 10x EV/EBITDA multiple is appropriate." — set `multiple_comparison` = {Future Outlook}.

- **Example (Future Outlook should be included):** "This multiple is justified by our 2025E earnings estimates, which reflect accelerating growth in cloud revenue." — set `multiple_comparison` = {Future Outlook}.
- **Example (Future Outlook should be included):** "We apply a 10-year revenue CAGR of 15%, terminal growth of 4%, and a 10% discount rate to derive our price target." — set `multiple_comparison` = {Future Outlook}.
- **Example (Future Outlook should not be included):** "Our price target implies a 10x EV/EBITDA multiple." — set `multiple_comparison` = {null}.
- **Example (Future Outlook should not be included):** "The implied CY11 PE is 20x." — set `multiple_comparison` = {null}.
- **Example (Future Outlook should not be included):** "This implies a multiple of approximately 31x our 2022E earnings per share (EPS) estimate, which is above the company's historical average." — set `multiple_comparison` = {Historical Market Values}.
- **Example (Future Outlook should not be included):** "Our $56/share price objective is based on a 19x multiple applied to our 2011 EPS estimate, which is a discount to its historical mean of 20x." — set `multiple_comparison` = {Historical Market Values}.

Appendix D List of Topics and Associated Categories

Table B1: Main topics This table presents all topics included in *Standardized Label 1* used in the prompt in Step 3 of Appendix B.

Category	Topic	Category	Topic	Category	Topic
Pricing	Pricing power	Corp. investment	Investment opportunities	Legal	Legal environment
Pricing	Pricing pressure	Corp. investment	Automation and robotics	Legal	Lawsuits and settlement
Pricing	Pricing strategy	Prod./Utilization	Operating leverage and scalability	Regulation	Regulatory environment and compliance
Product mkt. comp.	Barriers to entry	Prod./Utilization	Utilization	Regulation	Cybersecurity and data protection
Product mkt. comp.	Collusion	Financing	Credit rating	Regulation	Regulatory approval
Product mkt. comp.	Cost structure and efficiency	Financing	Credit availability and access	Domestic gov.	Political environment
Product mkt. comp.	Customer concentration	Financing	Debt level/leverage	Domestic gov.	Fiscal environment
Product mkt. comp.	Customer loyalty and retention	Financing	Cash and liquidity	Domestic gov.	Government debt
Product mkt. comp.	Economies of scale	Financing	Stock offerings and issuances	Domestic gov.	Taxation
Product mkt. comp.	Market share	Financing	PE/VC/alternative financing	Domestic gov.	Subsidies
Product mkt. comp.	Market size	Financing	Risk management/hedging	Domestic gov.	Government procurement
Product mkt. comp.	Monopoly	Payout	Stock buybacks and repurchases	Domestic gov.	Immigration
Product mkt. comp.	New entrants	Payout	Dividend policy and strategy	Domestic gov.	Infrastructure
Product mkt. comp.	Oligopoly	M&A	Spin-off and divestitures	Housing market	Housing supply
Product mkt. comp.	Threat of substitute	M&A	Synergy and integration	Housing market	Mortgage rates
Product mkt. comp.	Supplier concentration	M&A	Mergers and acquisitions (M&A)	Housing market	Housing demand
Product mkt. comp.	Vertical integration	M&A	Partnership/Joint-venture	Housing market	Home prices
Product mkt. comp.	Market expansion (industry)	M&A	Restructuring	Demographic	Demographic changes
Product mkt. comp.	Market expansion (geography)	M&A	Antitrust	Demographic	Income inequality
Customer	Customer acquisition	Contracting	Licensing and royalty	Crisis	Natural disasters
Customer	Customer demand/spending	Contracting	Contract terms and duration	Crisis	Pandemic and public health issues
Brand	Brand equity and reputation	Contracting	Covenants	Crisis	Wars and terrorism
Brand	Marketing and advertising	Governance	CEO/Senior management	Crisis	Economic sanctions
Product	Product/service mix	Governance	Management guidance	Raw input	Commodity/raw material prices
Product	Product/service quality	Governance	Corporate governance	Raw input	Oil and gas prices
Product	Product/service positioning and different.	Governance	Corporate ownership structure	GDP growth	Gross domestic product (GDP)
Product	Product life cycle and obsolescence	Labor market	Corporate social responsibility (CSR)	GDP growth	Economic recovery
Product	New products/services	Labor market	Employee safety	GDP growth	Economic recession
Product	Product recall	Labor market	Hiring and layoff	Inflation	Inflation and CPI
Supply chain	Production capacity and planning	Labor market	Labor strikes	Foreign exchange	Foreign exchange and currency rates
Supply chain	Transportation and logistics	Labor market	Pensions	International	International trades/import and exports
Supply chain	Inventory management	Labor market	Employee satisfaction	International	Globalization
Supply chain	Order fulfillment/backlog/backorder	Labor market	Corporate culture	International	Tariffs
Supply chain	Supply chain agility and flexibility	Labor market	Outsourcing	International	Geopolitical environment
Supply chain	Supplier relationships and bargaining power	Labor market	Wages	Industry	Industry life cycle
Innovation / R&D	Products pipeline	Labor market	Unions	Discount rate	Discount rates/WACC/Cost of capital
Innovation / R&D	Innovation and R&D	Labor market	Unemployment rates	Systematic risk	Risk premium/Systematic risk
Innovation / R&D	Clinical trials	Labor market	Labor force participation	Interest rate	Interest rate/Treasury yield
Innovation / R&D	Trade secrets and patents	Labor market	Skills shortages	Interest rate	Monetary policy
Innovation / R&D	Intellectual property (IP)	Environment	Environmental policies and ESG	Valuation	Undervalued
Innovation / R&D	AI and cognitive applications	Environment	Pollution and toxic waste	Valuation	Overvalued
Innovation / R&D	Technology lifecycle	Environment	Weather and climate change	Valuation	Fair-valued
Corp. investment	Capital expenditure (Capex)	Environment	Energy transition		

Table B2: Alternative topics This table presents all topics included in *Standardized Label 2* used in the prompt in Step 3 of Appendix B.

Category	Topic	Category	Topic	Category	Topic
Undetermined	Profitability Margins	Undetermined	Cash Flows	Undetermined	Discounted Cash Flow (DCF)
Undetermined	Sales And Revenues	Undetermined	Industry Trends	Undetermined	Sum-Of-The-Parts (SOP)
Undetermined	Costs And Expenses	Undetermined	Macroeconomic Trends	Undetermined	Other
Undetermined	Earnings, EBITDA, And EPS	Undetermined	Valuation Multiples		

Appendix E Equity Report Excerpt and LLM Output

Excerpt (Deutsche Bank, The Walt Disney Company, 2017-02-09): "**Theme Parks performance remains strong**, with **cost efficiencies in the US Parks** coming through as we expected. Mgt seems to be **planning to increase prices in the domestic parks**, or expand demand-based pricing to multi-day tickets. They hinted at some pricing action, without specificity. **International Parks' performance improved**, generating OI growth."

Output structure:

- {Theme park performance}//{firm}//{earnings/cash flow: earnings}//{positive}//{strong performance}//{present}//{**Theme park performance remains strong**}
- {Cost efficiencies}//{firm}//{earnings/cash flow: costs}//{positive}//{cost savings}//{present}//{**cost efficiencies in the US**}
- {Park pricing}//{firm}//{earnings/cash flow: sales}//{positive}//{price increase}//{near-future}//{**planning to increase prices in**}
- {International parks performance}//{firm}//{earnings/cash flow: earnings}//{positive}//{improved results}//{past}//{**International Parks' performance improved**}

Appendix F Multiples Reference Points

Firm Performance as a Reference Point

Forward-Looking Interpretation of Historical Data

1. **Equita, LUX.MI^C19, 2004-02-03:**
   *"[**We apply a**] P/E of 20.5x vs. 22.3x historical average in the last 5 years, despite the potential synergies that can be achieved in the following years."*
2. **UBS Equities, URI.N, 2014-01-21:**
   *"The 6.1x [**we use**] is within the historical range, which we think is supported by improvement in the business (execution and mix have driven margins higher, and cash flow and returns are improving)."*
3. **Jefferies, TXN.OQ, 2020-02-05:**
   *"A 24x P/E [**used in our valuation**] is in the top half of the 5-yr range of 16x–25x and we think warranted at the bottom of inventory correction and given its exposure to secular trends (IoT, Auto, Industrial, Consolidation)."*
4. **Morgan Stanley, EIGE.L^C20, 2008-02-27:**
   "P/E Take the market 2010 P/E of 10.5x and assume a 20% discount for below-market long-term growth."

Backward-Looking Interpretation of Historical Data

1. **Morgan Stanley, PII.N, 2021-03-19:**
   "This is in line with Polaris' historical multiple of 16.4x, which we think is warranted given the opportunity ahead."
2. **UBS Equities, HES.N, 2013-04-24:**
   "$75 PT assumes 4.8x normalized 2013E DACF, in-line with its historical average."
3. **Deutsche Bank, GLW.N, 2023-04-26:**
   "P/E valuation ($35, using 18x 2023E core EPS, in line with the historical average)."
4. **Kepler Chevreux, BDTG.DE, 2023-01-13:**
   "At our TP [target price], BDT would be trading in line with the average of its five-year EV/EBITDA multiple (6x)."

Peer Group and Industry as a Reference Point

1. **UBS, BBVA.MC, 2009-03-04:**

   "We believe Spirent is cheap at current levels and based on our forecasts Spirent trades on 7x 2010E P/E and an EV/sales of 0.8x, compared to Agilent on 9.4x and 0.9x respectively."

2. **Bank of America, TT.N, 2017-02-03:**

   "Our PO of $85 is based on applying a 17x P/E to our '18 EPS forecast, broadly in line with the average for the peer group."

3. **Nomura, SCC.BK, 2013-01-03:**

   "We assign a multiple of 10x, in line with its regional chemical peers."

4. **Credit Suisse, PER.L^L14, 2014-03-14:**

   "The target price is derived by putting Perform on 28.5x, a 20% discount to internet peers (on 35.6x 2014e)."

Appendix G Appendix Figures and Tables

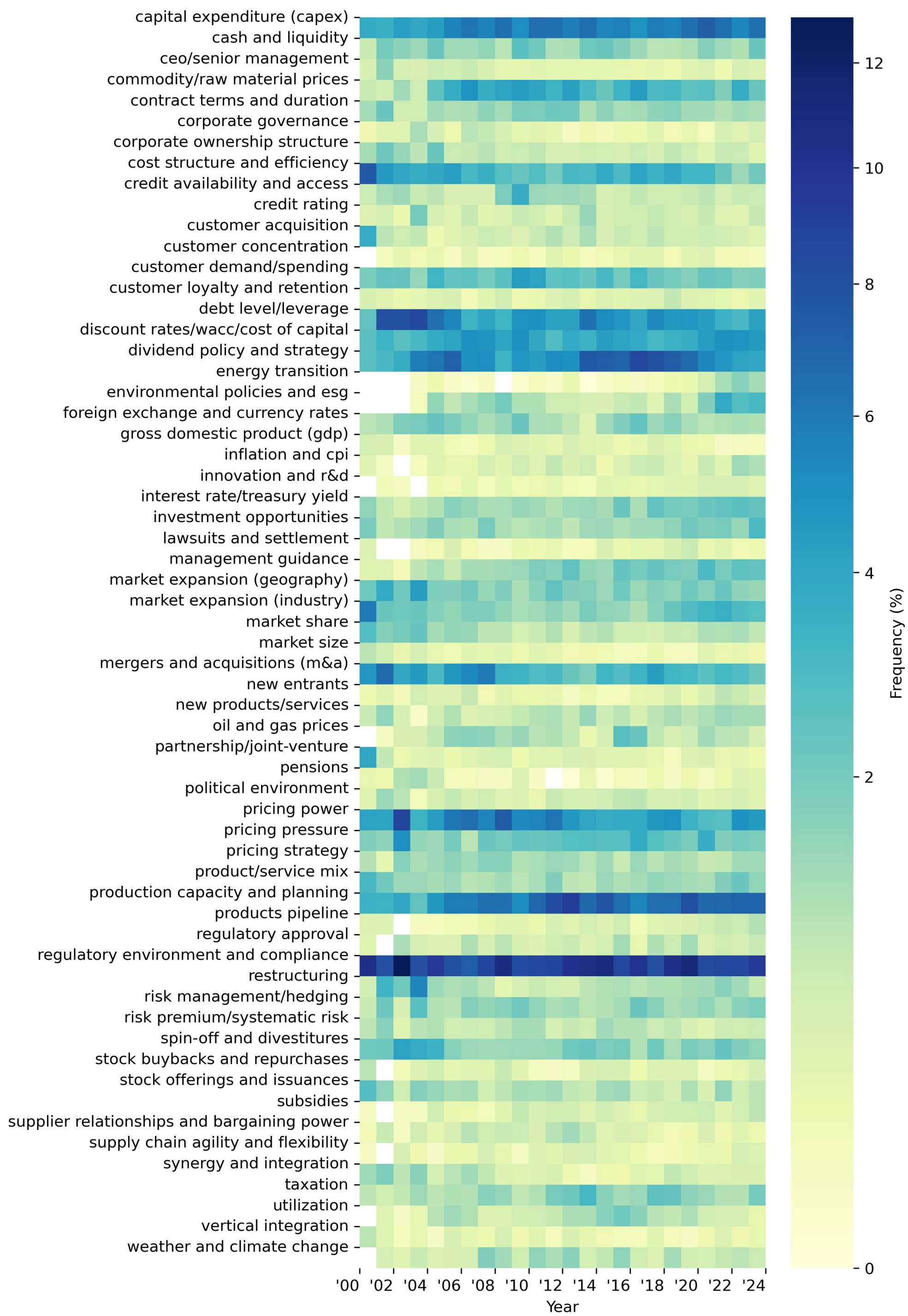


Figure E.1:
Topic Attention Allocation Over Time – Utilities

This figure plots attention allocation to the 60 most frequently discussed across utility firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

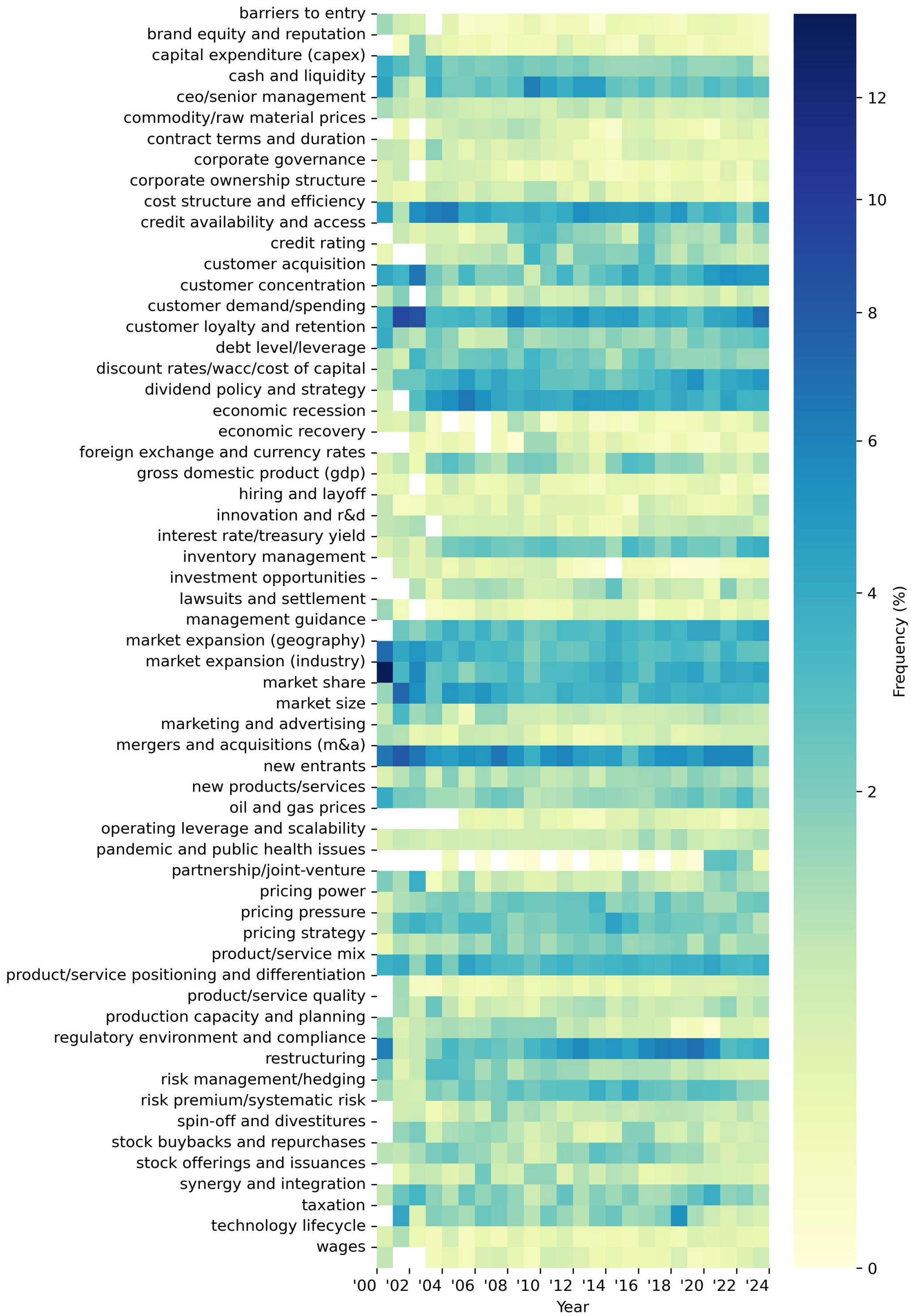


Figure E.2:
Topic Attention Allocation Over Time – Finance and Insurance

This figure plots attention allocation to the 60 most frequently discussed across finance and insurance firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

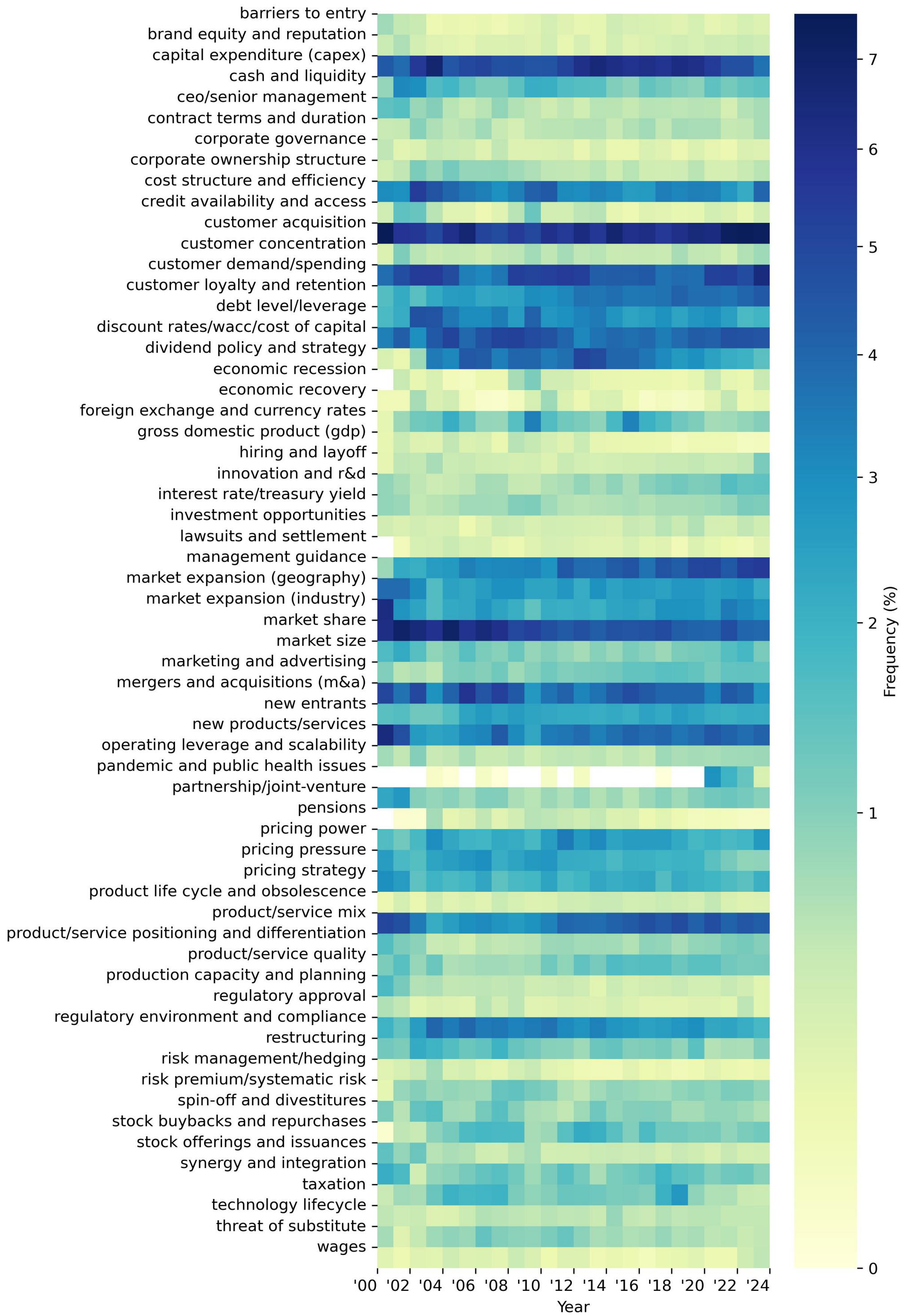


Figure E.3:
Topic Attention Allocation Over Time – Information

This figure plots attention allocation to the 60 most frequently discussed across information firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

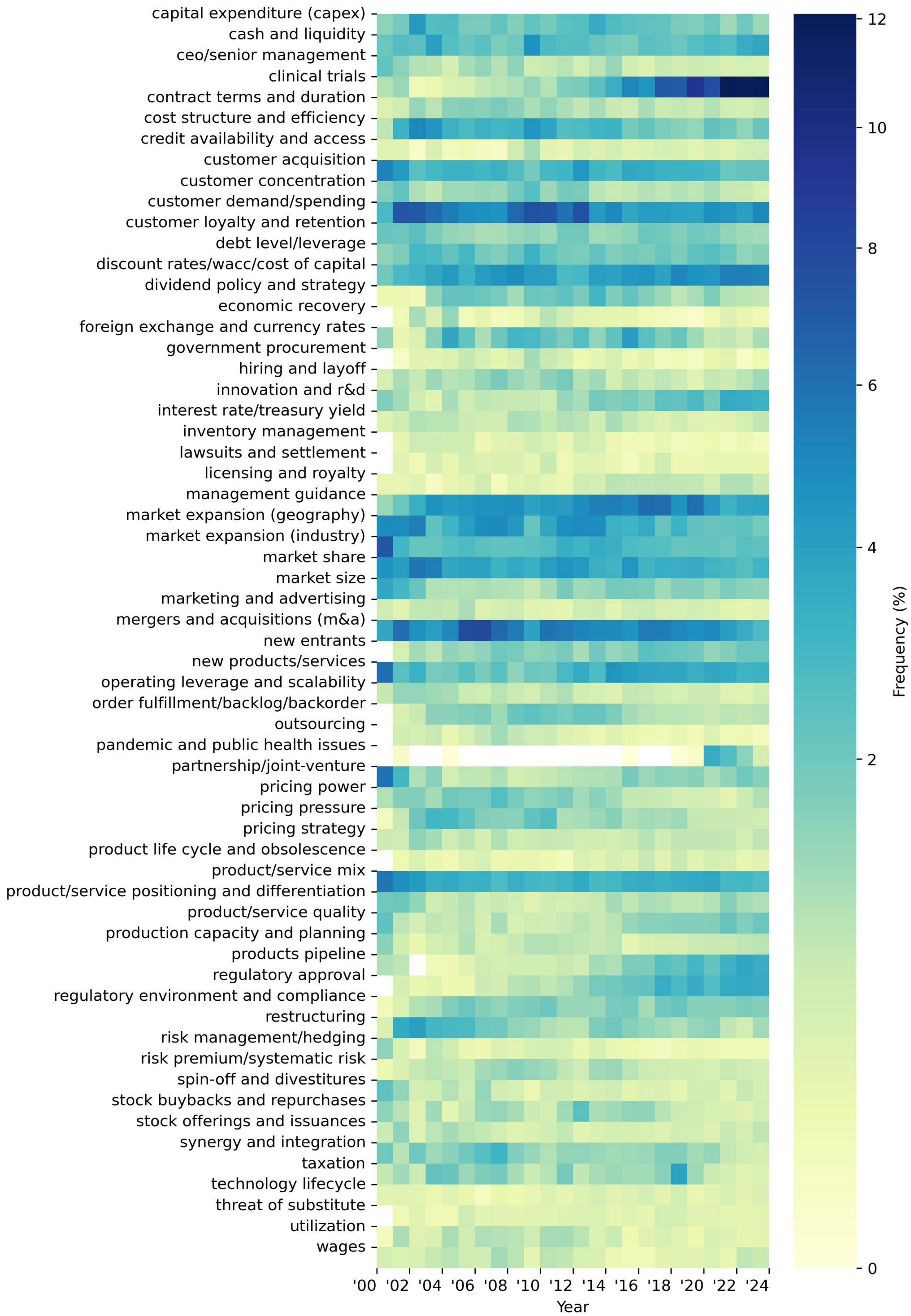


Figure E.4:
Topic Attention Allocation Over Time – Professional, Scientific, and Technical Services

This figure plots attention allocation to the 60 most frequently discussed across professional services firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

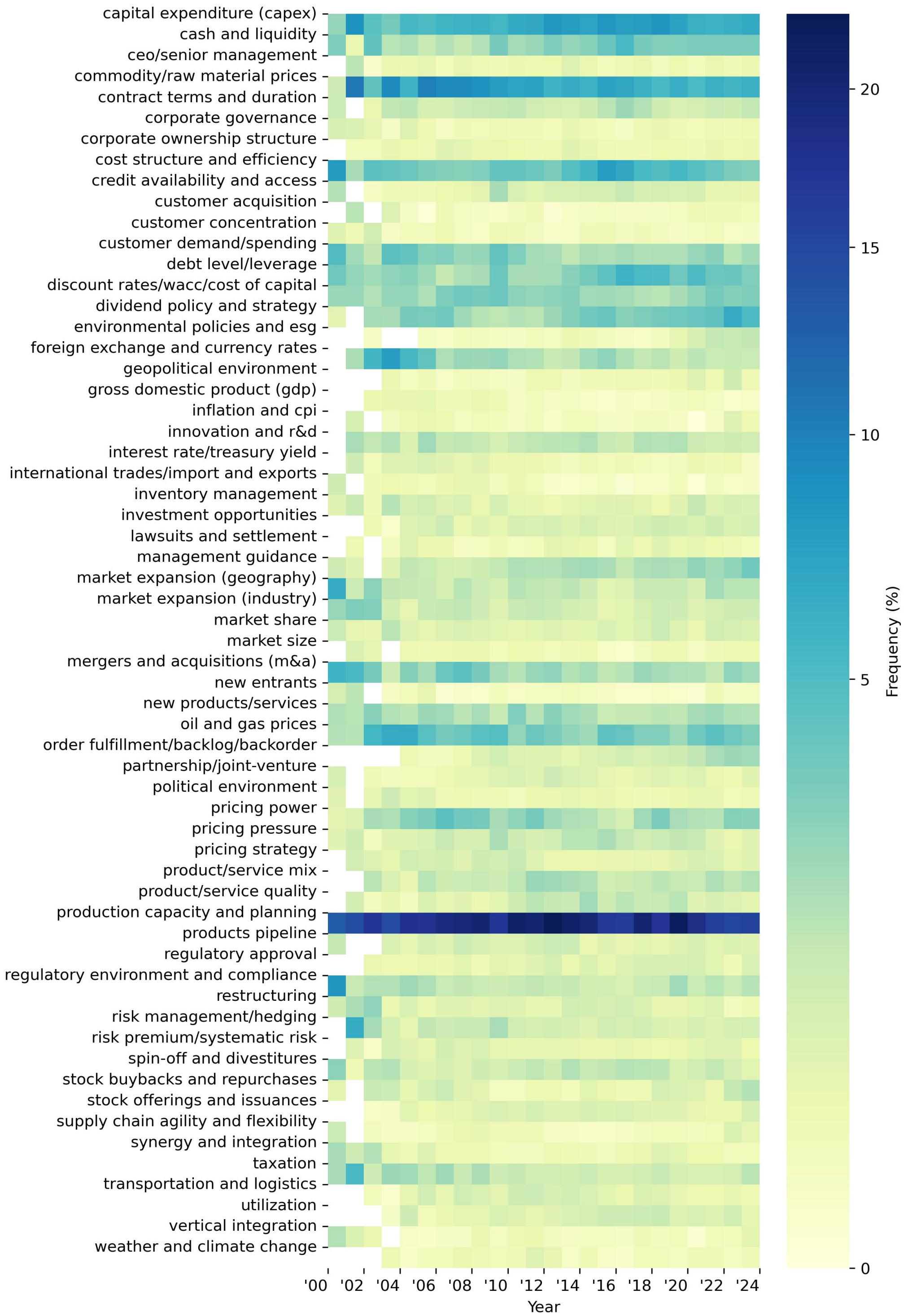


Figure E.5:
Topic Attention Allocation Over Time – Mining / Oil & Gas

This figure plots attention allocation to the 60 most frequently discussed across mining and oil & gas firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

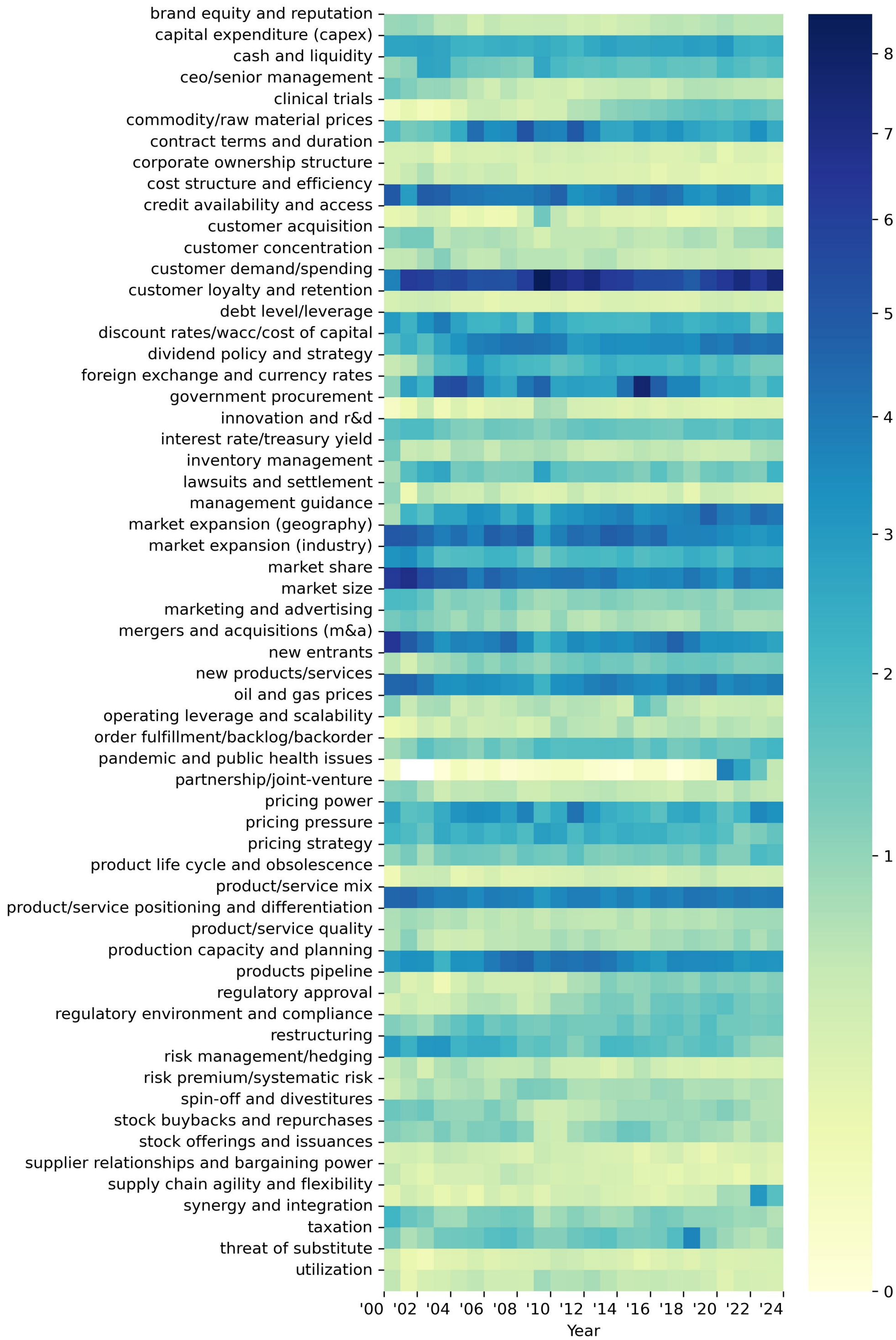


Figure E.6:
Topic Attention Allocation Over Time – Manufacturing

This figure plots attention allocation to the 60 most frequently discussed across manufacturing firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

Figure E.7:
Topic Attention Allocation Over Time – Retail Trade

This figure plots attention allocation to the 60 most frequently discussed across retail trade firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

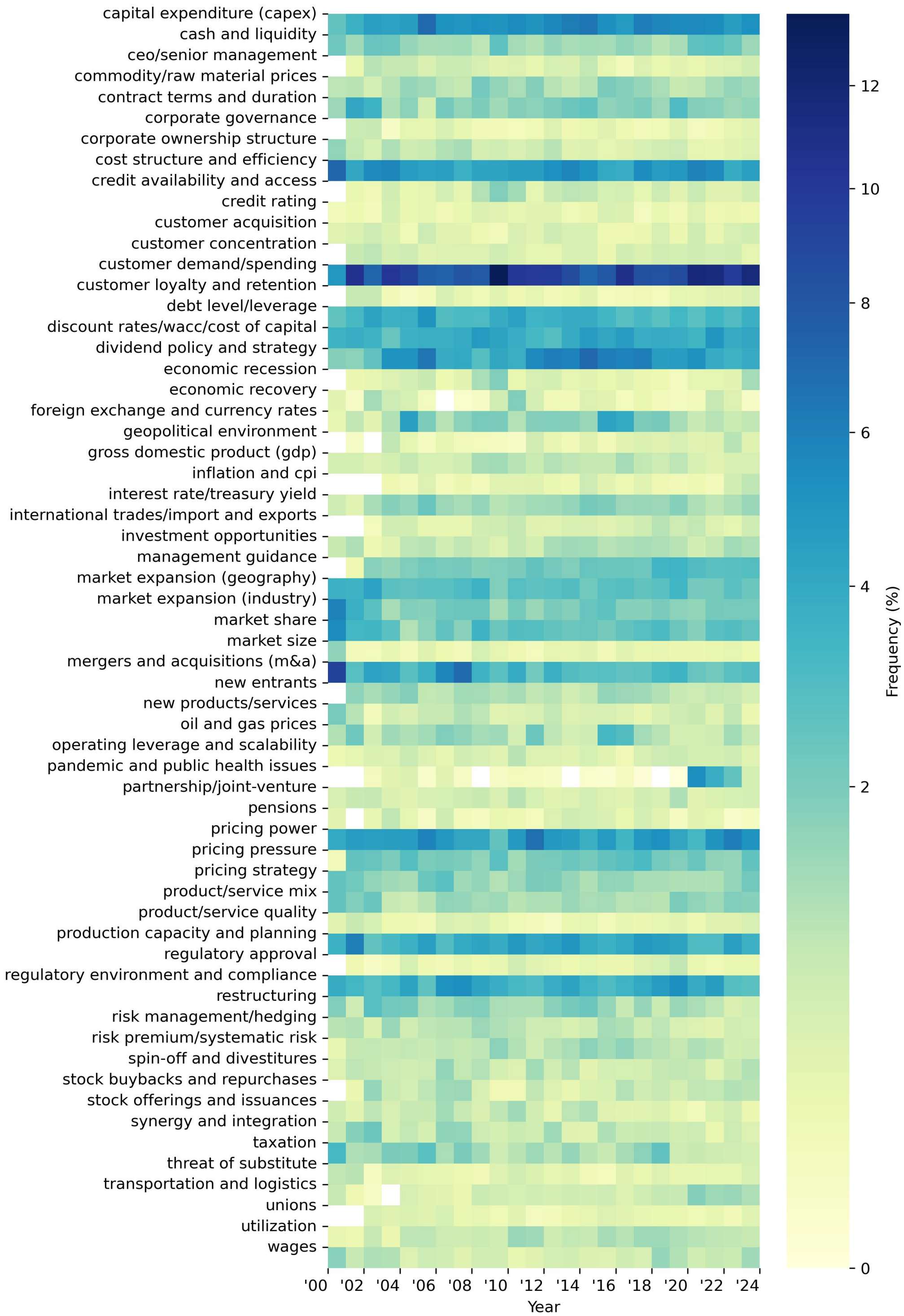


Figure E.8:
Topic Attention Allocation Over Time – Transportation

This figure plots attention allocation to the 60 most frequently discussed across transportation firms in our sample from 2000 to 2023. The x-axis denotes the year in which reports were produced, and the y-axis lists the top 60 topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

Figure E.9:
Topic Attention Allocation Over Time – Entities

This figure plots the attention allocation to the 20 most important topics for the three types of entities we focus on in the sample–firm, industry, and macroeconomy–for the period 2000–2023. The first block denotes firm topics. The second block denotes industry topics. The third block denote the macroeconomy topics. The x-axis denotes the year in which reports were produced, and the y-axis lists the topics. The color-coded heatmap is shown on the right, with yellow shades indicating lower values and dark blue shades representing the highest values in the sample. A non-linear gamma scale is used to enhance the visibility of variation across the full range of intensities. Frequency (in percent) corresponds to the share of the average report in a given year that discusses each topic. For example, a value of 5% for topic A in 2005 indicates that, on average, 5% of the content in reports written that year focused on topic A.

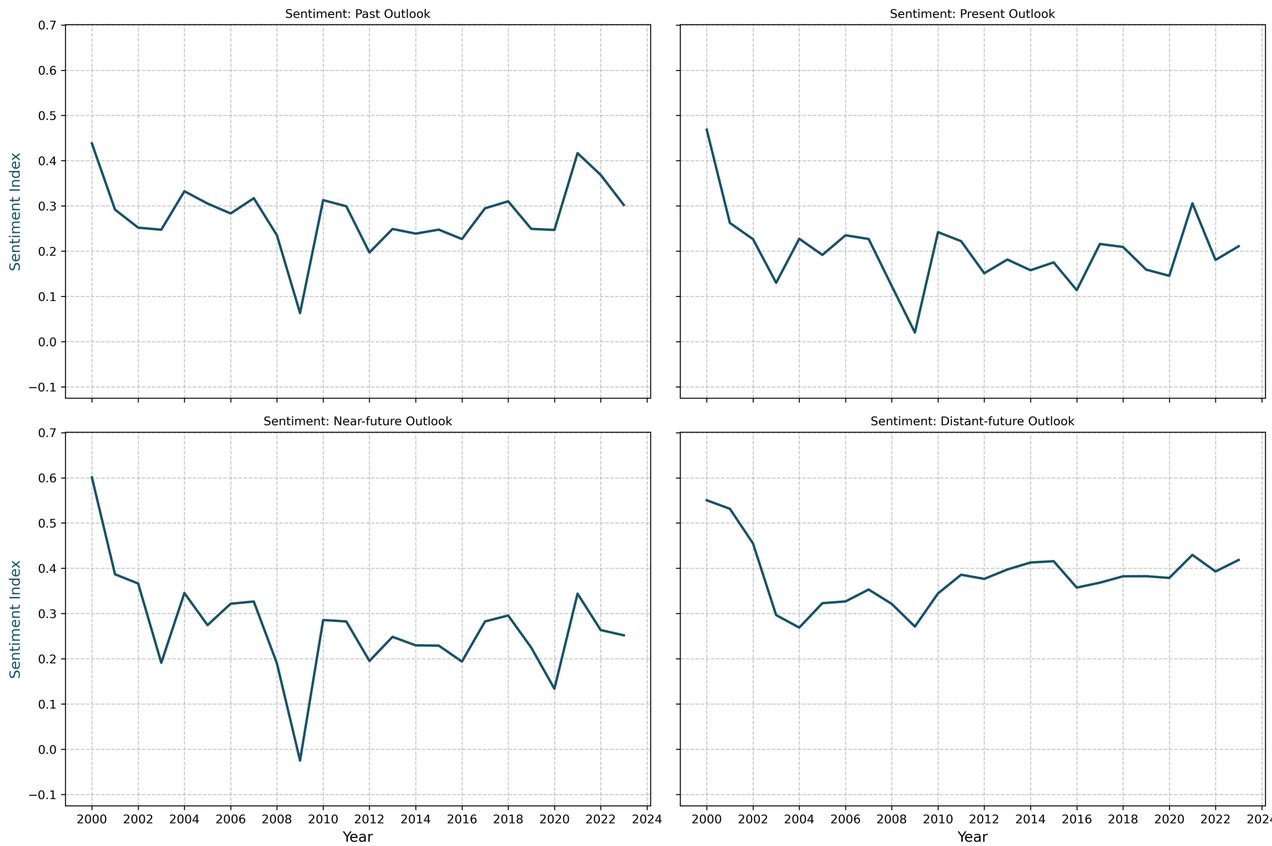


Figure E.10:
Average Sentiment by Time Outlook

This figure plots average sentiment over time by time outlook category—past, present, near future (1–3 years), and distant future (>3 years)—for the period 2000–2023. The x-axis denotes years. The y-axis represents the average sentiment in reports written in a given year, measured on a scale from -1 (negative) to $+1$ (positive), with 0 indicating neutral sentiment.

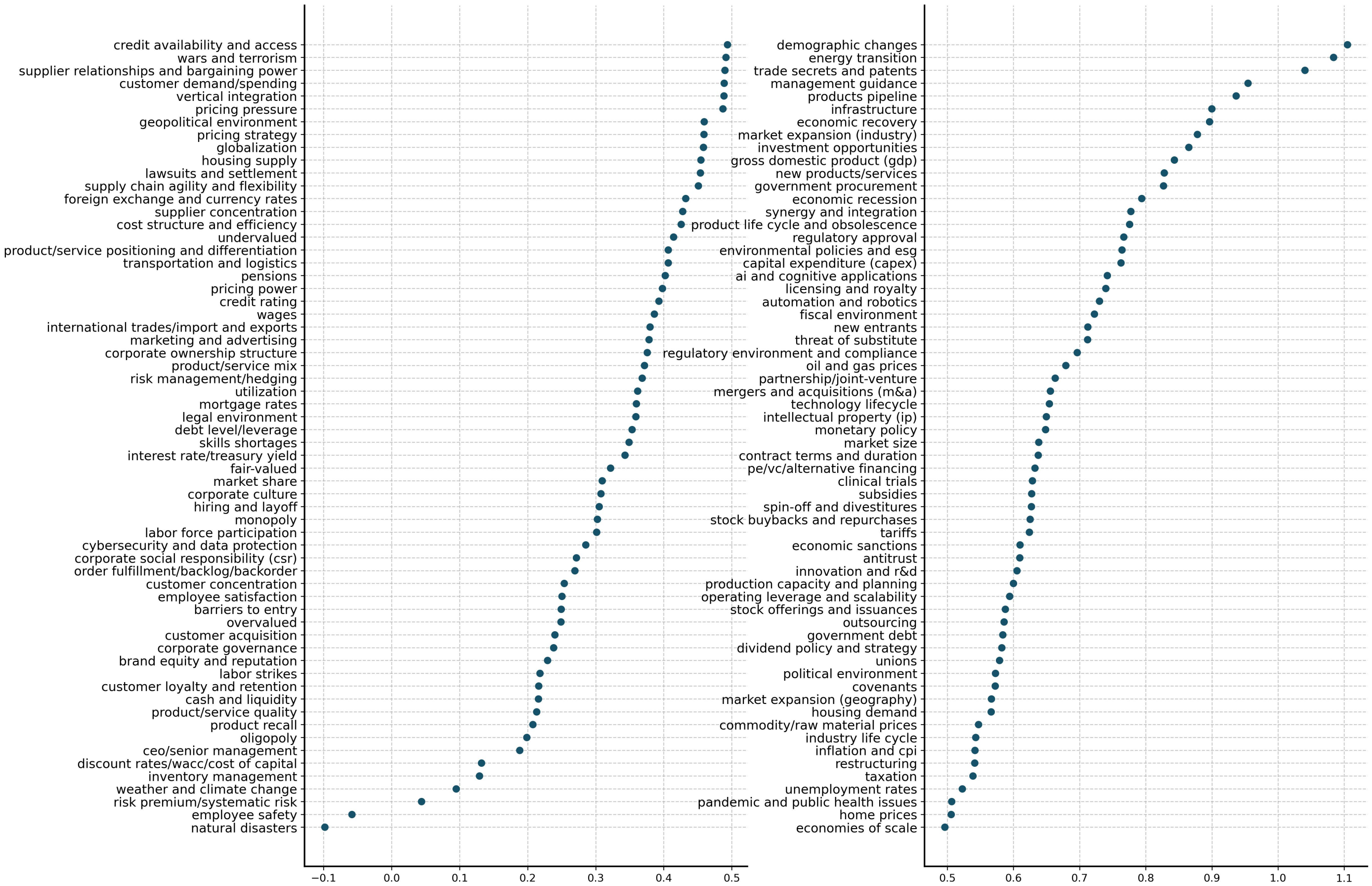


Figure E.11: Topic Outlook

This figure plots the average topic outlook associated with the most common topics in our sample over the period 2000–2023. The x-axis represents the average outlook for each topic, calculated on a scale of −1 (past), 0 (present), and +1 (near-future), and +2 (distant-future). The y-axis displays the corresponding topic labels.

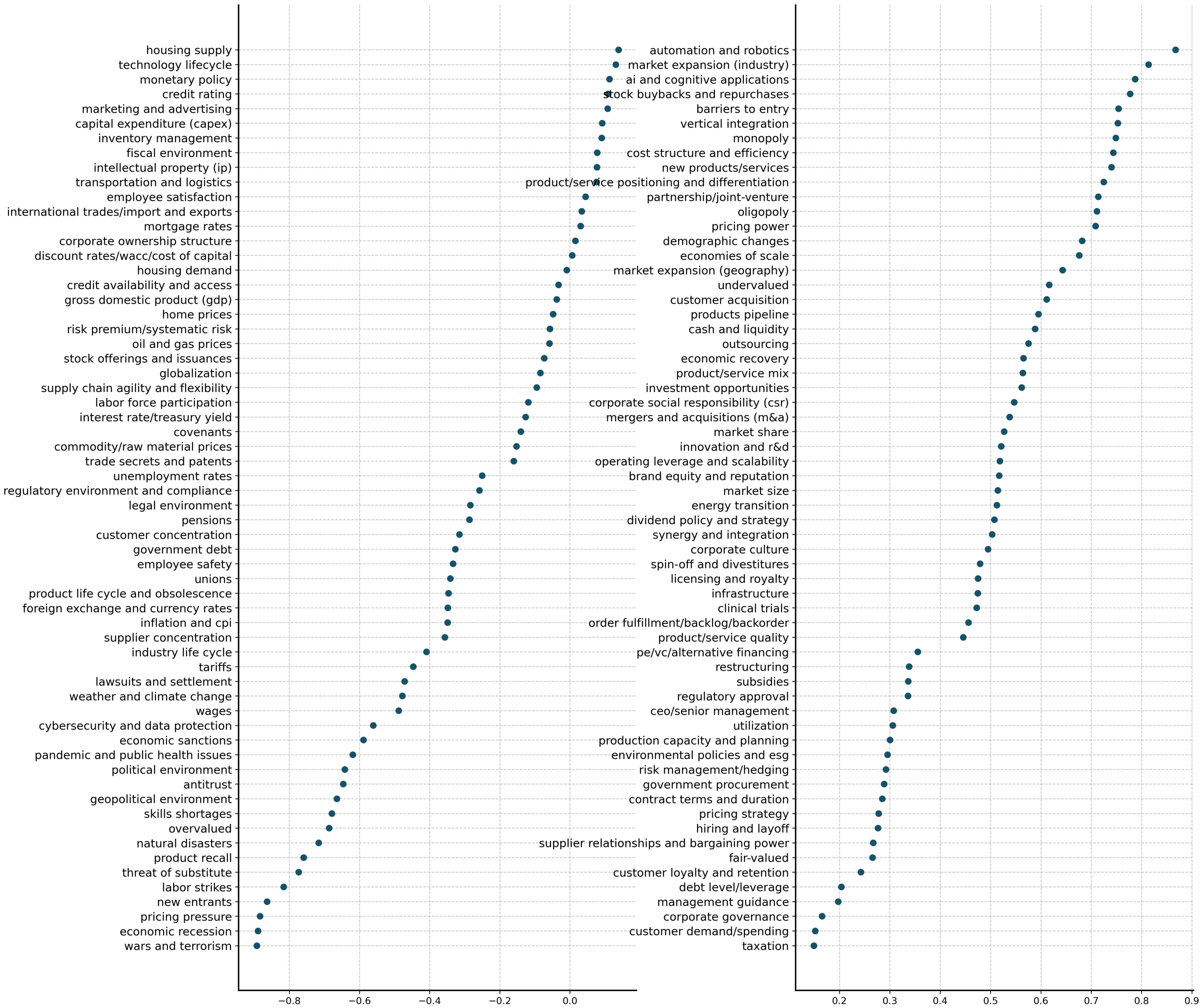
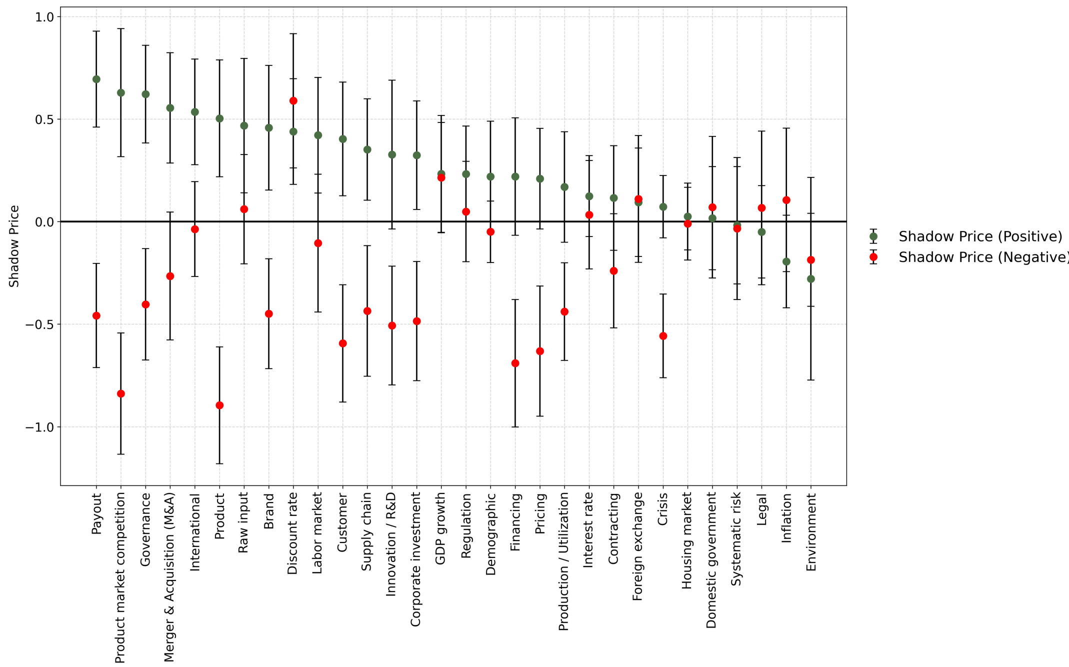


Figure E.12: Topic Sentiment

This figure plots the average sentiment associated with the most common topics in our sample over the period 2000–2023. The x-axis represents the average sentiment for each topic, calculated on a scale of −1 (negative), 0 (neutral), and +1 (positive). The y-axis displays the corresponding topic labels.

(A) Shadow Price.



(B) Shadow Price Delta between Analyst Pairs.

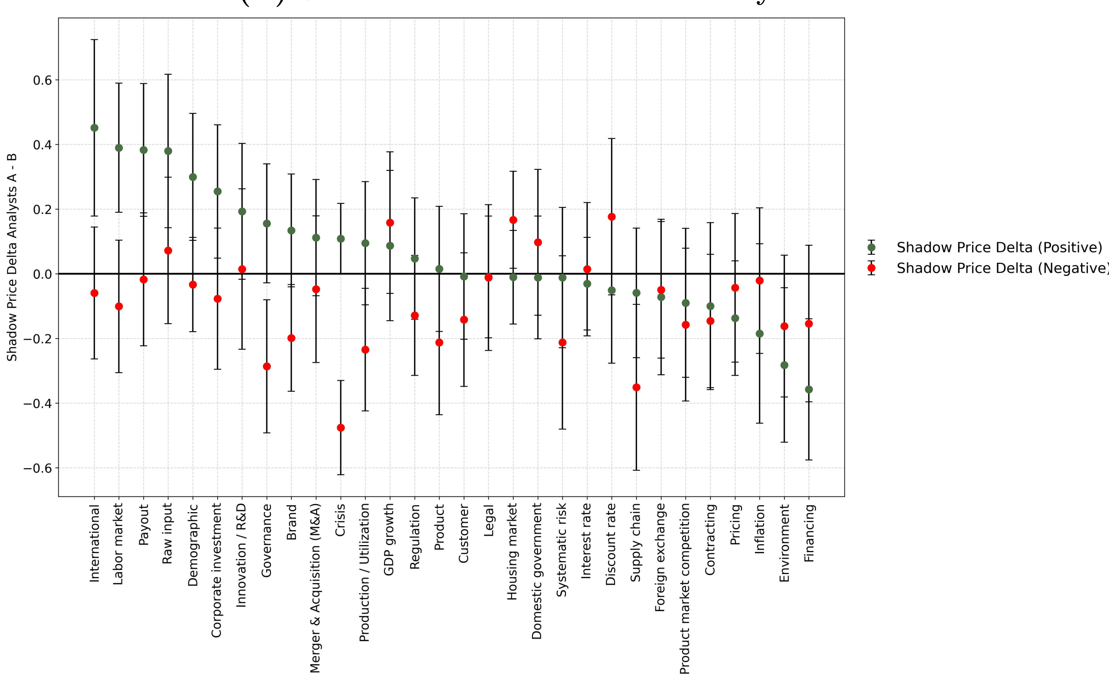


Figure E.13:
Shadow Price Estimates

This figure plots the shadow prices associated with each of the aggregate categories in our sample (see Appendix B). In Panel A, the green and red dots represent estimates of $\lambda^+_{k,B}$ and $\lambda^-_{k,B}$, respectively. In panel B, the green and red dots represent estimates of $\lambda^+_{k,A} - \lambda^+_{k,B}$ and $\lambda^-_{k,A} - \lambda^-_{k,B}$, respectively. For both cases, the coefficients are jointly estimated from the regression $\frac{\mathbb{E}^i_t[p_{ft}] - \mathbb{E}^j_t[p_{ft}]}{\mathbb{E}^i_t[p_{ft}] + \mathbb{E}^j_t[p_{ft}]} = \sum_{k=1}^{K} \lambda^+_k (a_{ikft} - a_{jkft})^+ + \lambda^-_k (a_{ikft} - a_{jkft})^- + (\lambda^+_k - \lambda^+_k)\, a^+_{jkft} + (\lambda^-_k - \lambda^-_k)\, a^-_{jkft} + \epsilon_{ft}$ where $a^+_{k,i,j,t}$ ($a^-_{k,i,j,t}$) denotes the fraction of analyst i's report on firm j in year t allocated to topic k when discussed positively (negatively), and $\lambda^+_{k,i}$ ($\lambda^-_{k,i}$) are the associated shadow prices used by analyst i to price topic k scaled by the corresponding signal. The "undetermined" and "valuation" categories are excluded from the regression exercise. 90% confidence interval are measured using heteroskedastic consistent standard errors clustered at the firm level. $a^+_{k,i,j,t}$ and $a^-_{k,i,j,t}$ are scaled by their in-sample standard deviation to offer a direct measure of comparison across topics and topic sentiment direction.

Table E.1:
Comprehensiveness of Extracted Arguments and Topics

This table presents results examining the comprehensiveness of extracted arguments and topics, focusing on narratives related to cash-flow channels: sales, costs, earnings, and margins. Panel A reports the proportion of topic-report observations associated with one, two, three, or four cash-flow channels. Panel B reports the proportion of topic-report observations associated with one, two, or three distinct sentiments.

Panel A: Number of Cash-flow Channels per Topic-Report	**Share of Sample**
1	86.85%
2	11.78%
3	1.26%
4	0.11%
Panel B: Number of Sentiment per Topic-Report	**Share of Sample**
1	76.8%
2	19.2%
3	3.9%

Table E.2:
Informativeness of Extracted Information by Report Location – Sentiment

This table presents results examining the informativeness of extracted information for analysts' 12-month price target forecasts across different sections of the report. $Sentiment^{\text{Quintile}}_{i,j,t}$ denotes the sentiment associated with the portion of the report falling within a given quintile of report length. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedasticity-consistent estimators clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	$\ln(\frac{E_{i,j,t}[P_{j,t+1}]}{EPS_{j,t}})$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(β_1) Sentiment$^{0-20^{th}pct}_{i,j,t}$	0.06***					0.05***	0.05***	0.05***	0.05***
	(0.00)					(0.00)	(0.00)	(0.01)	(0.01)
(β_2) Sentiment$^{21-40^{th}pct}_{i,j,t}$		0.04***				0.02***	0.02***	0.02***	0.02***
		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
(β_3) Sentiment$^{41-60^{th}pct}_{i,j,t}$			0.04***				0.02***	0.02***	0.01***
			(0.00)				(0.00)	(0.00)	(0.00)
(β_4) Sentiment$^{61-80^{th}pct}_{i,j,t}$				0.04***				0.02***	0.02***
				(0.00)				(0.00)	(0.00)
(β_5) Sentiment$^{81-100^{th}pct}_{i,j,t}$					0.01***				0.00
					(0.00)				(0.00)
Firm*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37,299	37,043	36,198	35,286	35,101	36,661	35,257	33,344	31,615
F Statistics	221.832	150.799	116.885	112.657	14.638	128.435	85.726	66.791	52.534
R^2	0.944	0.944	0.943	0.945	0.943	0.944	0.944	0.946	0.945

Table E.3:

Mental Model Alignment Across Analysts – Outlook, Valuation Channels, and Sentiment

This table presents results relating sentiment, time outlook, and valuation channels. The analysis is conducted at the analyst pair (A, B), topic k, firm j and year t level. The dependent variable, *Same sentiment*, equals one if both analysts share the same sentiment for the overlapping topic k and zero otherwise. *Same outlook* equals one if both analysts share the same outlook to topic k and zero otherwise. *Same valuation channel* equals one if both analysts assign the same valuation channel to topic k and zero otherwise. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedastic-consistent estimators and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	Same sentiment$_{A,B,k,j,t}$			
	(1)	(2)	(3)	(4)
(β_1) Same outlook$_{A,B,k,j,t}$	0.03***		0.02***	0.02***
	(0.00)		(0.00)	(0.00)
(β_2) Same valuation channel$_{A,B,k,j,t}$		0.03***	0.03***	0.02***
		(0.00)	(0.00)	(0.00)
Topic*Firm*Contributor$_A$*Contributor$_B$ FE	No	No	No	Yes
Topic*Firm*Year FE	No	No	No	Yes
Observations	3,672,802	3,672,802	3,672,802	3,469,800
F Statistics	1174.05	760.75	1015.16	619.89
R^2	0.00	0.00	0.00	0.25

Table E.4:
Semantic Similarity

This table provides an example of the semantic similarity between the topic "Antitrust" and the other topics included in the analysis. A higher value of semantic similarity indicates a greater likelihood that when analysts discuss these two topics in their analysis, they are effectively referring to a similar concept. We measure semantic similarity in two steps. First, for each topic used in the analysis, we generate the associated embedding vector using OpenAI embedding API "text-embedding-ada-002." In a second step, we measure the Cosine similarity for each pair of topic. In unreported analysis, we ensure the robustness of our results and recomputed the semantic similarity using Python's *spacy* package to generate the topics' embedding. Any results derived from those embeddings are robust to such alternative.

Topic	**Semantic similarity with "Antitrust"**
Collusion	86%
Oligopoly	85%
Barriers to entry	82%
Vertical integration	82%
[...]	
Market share	81%
Partnership / Joint-venture	81%
Mergers and acquisitions (M&A)	80%
Intellectual property (ip)	80%
[...]	
Home prices	78%
Monetary policy	78%
Income inequality	78%
Corporate social responsibility (csr)	78%
[...]	
Capital expenditure (capex)	74%
Discount rates / WACC / Cost of capital	73%
Debt level / Leverage	72%
Risk premium / Systematic risk	72%

Table E.5:
Topic Selection

This table presents results on mental model completeness. The analysis is conducted using an Elastic Net estimator with cross-fold validation. The table reports the number of topics selected by the Elastic Net as statistically relevant for explaining the dependent variable, out of the total number of topics considered. The estimation in Panel A is performed on the longest reports of the sample (top 10 percent of report length). The estimation in Panel B is performed on the full sample. In Column 1 of Panel A, we implement the Elastic Net procedure on the vectors of coefficients omitted topics with associated with a positive sentiment ($\beta^{+}_{\text{Omitted}}$) and those with a negative/neutral sentiment ($\beta^{-}_{\text{Omitted}}$) controlling for the topics already discussed in the report in the following relation:

$$\ln(\frac{E_{i,j,t}[P_{i,t+1}]}{EPS_{j,t+1}}) = X^{+/Included}_{i,j,t}\phi^{+} + X^{-/Included}_{i,j,t}\phi^{-} + X^{+/Omitted}_{i,j,t}\beta^{+} + X^{-/Omitted}_{i,j,t}\beta^{-}$$

where $X^{+/Included}_{i,j,t}$ ($X^{-/Included}_{i,j,t}$) is a vector of binary variables each equal to 1 if the topic is discussed with a positive (negative/neutral) sentiment, and 0 otherwise, and $X^{+/Omitted}_{i,j,t}$ ($X^{-/Omitted}_{i,j,t}$) is a vector of binary variables each equal to 1 if the topic is omitted from the report but discussed by other analysts evaluating the same firm-year with a positive (negative/neutral) sentiment, and 0 otherwise. In Column 1 of Panel B, we estimate

$$\ln(\frac{E_{i,j,t}[P_{i,t+1}]}{EPS_{j,t+1}}) = X^{+/Included}_{i,j,t}\phi^{+} + X^{-/Included}_{i,j,t}\phi^{-}$$

In Column 2 of the panels, we estimate an elastic net feature selection on $\beta^{+}_{\text{Omitted}}$ and $\beta^{-}_{\text{Omitted}}$ controlling for the topics already discussed in the report in the following relation:

$$\frac{|E_{i,j,t}[P_{i,t+1}] - P_{i,t+1}|}{P_{j,t+1} + E_{i,j,t}[P_{i,t+1}]}) = X^{+/Included}_{i,j,t}\phi^{+} + X^{-/Included}_{i,j,t}\phi^{-} + X^{+/Omitted}_{i,j,t}\beta^{+} + X^{-/Omitted}_{i,j,t}\beta^{-}$$

where $\frac{|E_{i,j,t}[P_{i,t+1}]-P_{i,t+1}|}{P_{j,t+1}+E_{i,j,t}[P_{i,t+1}]}$ denotes the forecast error.

Panel A: Excluded topics	$\ln(\frac{E_{i,j,t}[P_{i,t+1}]}{EPS_{j,t}}$	$\frac{\lvert E_{i,j,t}[P_{i,t+1}]-P_{i,t+1}\rvert}{E_{i,j,t}[P_{i,t+1}]+P_{i,t+1}}$
	(1)	(2)
Number of excluded positive topic selected	0 / 139	10 / 139
Number of excluded negative/neutral topic selected	0 / 139	10 / 139
Observations	759	935
F Statistics	1.44	1.66
R^2	0.43	0.42
Panel B: Included topics	$\ln(\frac{E_{i,j,t}[P_{i,t+1}]}{EPS_{j,t}})$	$\frac{\lvert E_{i,j,t}[P_{i,t+1}]-P_{i,t+1}\rvert}{E_{i,j,t}[P_{i,t+1}]+P_{i,t+1}}$
	(1)	(2)
Number of included positive topic selected	139 / 139	139 / 139
Number of included negative/neutral topic selected	139 / 139	139 / 139
Observations	30,719	30,719
F Statistics	12.46	6.90
R^2	0.10	0.05

Table E.6:
Drivers of Topic Attention – The Case of Inflation Narratives

This table presents results examining the drivers of attention allocation to inflation. *Attention to Inflation* is defined as the share of analysts' reports allocated to discussing the topic "Inflation and CPI," at the analyst i, firm j, and year t level. *Inflation Sentiment* denotes the average sentiment toward inflation in a report, when the topic "Inflation and CPI" is included. $\text{Inflation}^{FirmHQCountry}$ and $\text{Inflation}^{AnalystCountry}$ refer to the realized inflation in the firm's HQ country and analyst country, respectively, in year t. $1^{\text{Inflation} > 2\%}$ is an indicator variable equal to one if inflation is above 2% and 0 otherwise. The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedastic-consistent estimators and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable:	Attention to $\text{Inflation}_{i,j,t}$					Inflation $\text{Sentiment}_{i,j,t}$	
Filter: Inflation in Firm HQ Country $\geq$ 2%:	N.A.	No	Yes	Yes	No	N.A.	No
Filter: Inflation in Analyst Country $\geq$ 2%:	N.A.	Yes	Yes	No	No	N.A.	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(β_1) $\text{Inflation}^{Analyst\ Country}_{c,t}$	0.04***	0.05*	0.06***	-0.01	-0.01	-0.03***	-0.06***
	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.01)	(0.00)
(β_2) $\text{Inflation}^{Firm\ HQ\ Country}_{p,t}$	0.02***	-0.00	0.03**	0.03**	-0.02	-0.00	0.00
	(0.01)	(0.03)	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)
Observations	18,946	4,152	6,172	2,334	6,288	975	651
F Statistics	57.50	2.39	54.93	3.86	1.77	6.63	91.71
R^2	0.02	0.00	0.03	0.01	0.00	0.01	0.03

Table E.7:
Drivers of Forecast Errors

This table reports results analyzing determinants of analyst forecast errors. The dependent variable, *Forecast error*, is defined as the absolute difference between the analyst's price target and the realized stock price one year after the forecast date, scaled by the sum of the price target and the realized price at that horizon. Observations are at the analyst-firm-year (i, j, t) level. *DCF Usage* is an indicator equal to 1 if the analyst's price target is derived from a discounted cash flow (DCF) model, and 0 otherwise. *Sentiment* measures the average sentiment expressed in the analyst's report. *Sales focus* represents the share of the report devoted to statements related to the firm's sales (instead of cost, margin, earnings/cash flow, and discount rate). The regressions are estimated using ordinary least squares. Standard errors (in parentheses) are estimated using heteroskedastic-consistent estimators and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Forecast Error Drivers	Forecast Error$_{i,j,t}$			
	(1)	(2)	(3)	(4)
(β_1) DCF usage$_{i,j,t}$	-0.01**	-0.01***		
	(0.01)	(0.00)		
(β_2) Sentiment$_{i,j,t}$			0.04***	
			(0.00)	
(β_3) Sales focus$_{i,j,t}$				0.02***
				(0.01)
Firm*Year FE	No	Yes	Yes	Yes
Observations	53,995	35,382	49,024	49,025
F Statistics	5.91	14.47	355.43	11.64
R^2	0.00	0.84	0.82	0.82