What Drives Very Long-Run Cash Flow Growth Expectations?*

Paul H. Décaire† and Marius Guenzel‡

February 13, 2025

Abstract

Using a novel dataset combining *terminal growth rate (TGR) expectations*, textual belief discussions, and demographics, we establish four findings about forecasters' truly long-run firm growth expectations. First, TGR expectations contain distinct information compared to shorter-horizon beliefs, correlate with firms' realized long-run growth, and are not contaminated by expected returns. Second, unlike for short-run expectations, forecaster characteristics explain a large share of the TGR belief variation. Third, belief heuristics, such as local extrapolation, become prevalent when forming TGR expectations for geographically and culturally distant firms. Lastly, divergent TGR expectations reflect subjective differences in emphasis on similar topics, not in topics considered.

JEL classification: D24, D25, D46, D84, G17, G31, G41

Keywords: Long-term expectations, behavioral finance, valuation, professional forecasters, analyst identities and backgrounds, local extrapolation

*For helpful comments and suggestions, we thank Aditya Chaudhry (discussant), Francesco D'Acunto (discussant), Kaushik Vasudevan (discussant), Spencer Kwon, and Laura Lindsey, as well as seminar and conference participants at the W. P. Carey School of Business at Arizona State University, the 2024 Mid-Atlantic Research Conference in Finance, the 2024 Behavioral Finance Conference at Bocconi, the 2024 CESifo Venice Summer Institute Expectation Formation, the 2024 Yale Junior Finance Conference, and the 2025 American Finance Association Annual Meeting. For their excellent research assistantship, we thank Eisa Alloughani, Sebastian Gortarez Ortiz, Tej Kulkarni, Aditeya Pattiwar, Andre Radic, Reed Simons, Cole Streitmatter, Mingyang Zhang, and Shuman Zhang. 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.

†W. P. Carey School of Business, Arizona State University, email: paul.decaire@asu.edu.

‡The Wharton School, University of Pennsylvania, email: mguenzel@wharton.upenn.edu.

1 Introduction

What shapes long-run growth expectations about firms? While abundant data on short-term beliefs has fueled extensive research into how market participants form short-horizon expectations, far less is known about longer-term outlooks. This gap is particularly salient since insights gained from studying short-term expectations may not translate well to long-term beliefs. Firms' current dynamics are often only loosely connected to their long-term outcomes. Further, more than 70% of firms' discounted cash flows accrue beyond a five- to ten-year horizon. Together, these factors underscore the centrality of long-run beliefs in valuations and suggest that determinants of short-run expectations may differ fundamentally from those at longer horizons—as Nagel (2021) notes, "collecting more data on long-run expectations would be useful" to refine models of belief formation.

This paper presents direct evidence on forecasters' *terminal growth rate (TGR) expectations*—the growth rate they expect firms to sustain in the very long run "until infinity"—organized around three high-level takeaways. First, TGR expectations embed distinct information relative to shorter-horizon forecasts, robustly correlate with firms' long-run growth (dominating shorter-run expectation series in horse race regressions), and are not confounded by investors' expected returns. Second, in contrast to short-run series, the cross-sectional variation in TGR beliefs is largely driven by forecasters' individual characteristics and demographics, with TGR forecasts being systematically influenced by, e.g., local extrapolation and cultural factors. Third, although forecasters consider similar factors when forming long-run projections, they assign differing weights to these factors, driving quantitative forecast heterogeneity. Broadly, our findings suggest that individual heterogeneity becomes increasingly important in *distant* economic contexts—whether temporal or geographical—making such cases more abstract and malleable to subjective experiences.

To conduct the study, we combine two datasets to obtain more than 41 thousand TGR forecasts, each associated with forecasters' qualitative text discussions of their beliefs as well as detailed forecaster demographic and other background information. The first dataset, containing the TGR expectations and text discussions, is assembled directly from original equity reports by professional forecasters (sell-side analysts), using a combination of PDF parsing functions, optical character recognition, and AI-based document processing programs. It covers forecasts for 8,000 firms located in 49 countries between 2000 and 2023, and 5,300 individual forecasters.

The second dataset provides comprehensive background information on forecasters. We use personal identifiable information extracted from the equity reports (full name, employer, and real time location) to merge the TGR data with the universe of individuals included on LinkedIn obtained through Revelio Labs, containing more than 750 million individuals. Doing so, we are able to collect detailed personal information, including gender, race, age, and educational background, for 2,565 individuals located in 46 countries.

In the first part of the paper, we establish the systematic differences of long-term TGR expectations from shorter-term expectation series, document their significant real economic information content, and show that they are not contaminated by investor expected returns.

First, TGR expectations differ sharply from the long-term growth (LTG) series in IBES, the focal longer-run measure in prior research, which is explicitly designed to reflect a three- to five-year forecasting horizon. The two series exhibit a noticeably weak correlation ($\rho = 0.07$), highlighting a substantial disconnect in their information content. This disconnect mirrors patterns in realized growth, where 5- or 10-year growth shows no correlation with shorter-run growth. Additionally, on average, LTG expectations exceed TGR by an average of 10 percentage points (pp), reflecting that LTG captures forecasters' medium-term expectations (Bordalo et al., 2024b), rather than truly long-run horizons for which forecasters generally expected firms' growth to converge toward a mix of industry and regional trends.

Furthermore, forecasters' average TGR expectations are strongly and positively associated with firms' future realized earnings growth. Importantly, the strong positive association persists after controlling for LTG medium-run expectations and across forecasting horizons. Specifically, medium-run expectations do significantly predict shorter-term realized earnings growth (1–5 years, coefficient of 0.32), but the estimated relation weakens dramatically at longer horizons (3–13 years, coefficient of 0.04). By contrast, TGR expectations consistently predict earnings growth at long-run horizons (1–5, 3-8, and 3–13 years, coefficients of 0.55, 0.46, and 0.40, respectively). Economically, a one percentage point increase in average TGR expectations is associated with a 0.40 percentage point rise in 10-year earnings growth, ten times the magnitude of the IBES LTG coefficient. Overall, the dominance of TGR expectations in the long-horizon predictive regressions highlights the limited ability of previously studied measures to capture the effective very long-run beliefs of financial market participants.

In addition to forecasting realized earnings growth, TGR expectations contain a firm life cycle component, further validating their economic substance. Average TGR beliefs across forecasters decline monotonically with firm age, consistent with forecasters initially facing uncertainty about whether young firms will achieve long-term success or encounter early financial distress. As firms mature and uncertainty reduces, TGR beliefs increasingly align with macroeconomic and industry trends, while firm-specific factors, such as default risk, hold greater influence for younger firms.

Finally, we provide evidence that TGR expectations are not contaminated by expected returns due to reverse-engineering from prices, thus contributing to the broader debate on the extent to which survey-based and self-reported expectations reflect "true" expectations (Nagel, 2024). In a nutshell, if investors and analysts price assets using a different discount rate and forecasters reverse-engineer growth expectations from current prices, the observed variation in reported beliefs would partly reflect differences in discount rates. Recent work provides some, albeit indirect, evidence to rule out

extreme cases of reverse-engineering, showing that *both* analysts and investors adjust, at least to some degree, both expectations about returns and cash flow growth when pricing assets (Golez, 2014; Hillenbrand and McCarthy, 2024; Pruitt, 2024; Chen et al., 2013; De La O and Myers, 2021; Decaire et al., 2023) and that both groups likely use time-varying discount rates that rely on common pricing models, such as the CAPM, for discount rates (Brav et al., 2005; Barber et al., 2016; Berk and van Binsbergen, 2016; Andrei et al., 2023; Decaire and Graham, 2024).

Our setting uniquely positions us to conduct two direct tests for assessing belief contamination by expected returns due to reverse-engineering, as we observe *all* components of forecasters' valuation models. First, building on the recent work by Bordalo et al. (2024a,b)[1] studying LTG expectations we show that TGR expectations are negatively correlated with firms' future realized returns and, moreover, remain so even after controlling for forecasters' *subjective* expected returns included in the same valuation report. Controlling for subjective expected returns directly accounts for and holds constant forecasters' discount rates, thereby isolating the portion of belief variation driven by other factors, such as subjective growth expectations. Second, we derive that, if contamination arises from reverse-engineering growth expectations from prices, subjective expected returns must exhibit offsetting errors with TGR expectations to balance the price equation, implying a testable prediction: subjective expected returns and growth expectations should be mechanically linked through these errors. We find no meaningful economic or statistical relationship, consistent with the idea that reverse-engineering-driven contamination of long-term growth expectations is unlikely to be widespread among forecasters.

In the second part of the paper, we examine the determinants of variation in long-run TGR expectations and how these determinants differ from those of shorter-run expectation series, thereby informing models of belief formation. In a first step, we compare the explanatory power of cross-sectional drivers between TGR and short-term (one-year-ahead) growth expectations, performing variance decompositions using three fundamental sets of fixed effects (FE)—firm FEs (i.e., FEs for each firm for which forecasts are made), individual FEs (i.e, FEs for each forecaster), and brokerage house FEs (i.e., FEs for each analyst employer). We document significant differences. For one, these three sets of fixed effects explain at least 45% of the TGR variation, but at most 39% of short-run expectations. Additionally, the relative importance of FEs differs markedly. For short-run expectations, firm fixed effects account for the largest share of the explained variation (averaging 74% across specifications). Time-invariant firm factors matter for TGR beliefs, as well, though

[1] These papers provide evidence that, among other things, changes in LTG are linked to multi-year earnings growth and returns, with LTG predicting returns even when controlling for valuation ratios. While this evidence limits the plausibility of risk-premium variables consistent with the data, Nagel (2024) concludes that it still "cannot unequivocally reject the concern that analysts' reported LTG inadvertently captures risk premia."

to a lesser extent, accounting for only 39% of the explained variation on average.[2] By contrast, individual forecaster FEs are substantially more important for long-run beliefs. On average, they account for only 22% of the explained variation for short-run beliefs, but for 56% for TGR beliefs.[3]

Motivated by the significant role of individual forecasters in driving the variation in long-run beliefs compared to short-run beliefs, we then examine two key individual-level dynamics: (i) the role of personal characteristics, backgrounds, and experiences in explaining heterogeneity in long-term belief formation dynamics across forecasters, and (ii) the role of differences in the qualitative topic discussion in equity reports, i.e., differential allocation of attention, in shaping TGR belief dispersion across forecasters.

Related to (i), we first show that forecaster characteristics and demographics explain 25–29% of the variation in forecaster fixed effects in our variance decomposition analysis, with key factors being forecasters' age, race, and geographic location. This significantly exceeds, e.g., the 3–10% of variation explained by demographics in Giglio et al. (2021) for wealthy U.S. retail investors.[4] This greater explanatory power likely stems from several factors, including having multiple observations per forecasters across multiple years and firms, which facilitates the fixed effect estimation, the increased diversity of the sample, including forecasters from multiple countries, which allows us to disentangle the effects of firm and forecaster countries, and the stickiness of long-term expectations, which makes a substantial part of the variation cross-sectional (Decaire and Graham, 2024).

Additionally, we explore the role of forecasters' local experiences in shaping long-term beliefs and find strong evidence of local extrapolation. Forecasters' TGR expectations for firms in *other* countries are strongly predicted by recent GDP growth in their own country, controlling for the firm's country GDP growth. This pattern is not merely driven by regional shocks affecting clusters of countries since the effect is concentrated among distant countries. Nor can it be explained by special linkages between forecaster and firm countries, since the results are robust to including firm-country*forecaster-country fixed effects. Instead, as predicted by heuristic-based underpinnings of extrapolation, forecasters are particularly prone to extrapolating from their local context and conditions when forming beliefs for geographically distant firms that are economically disconnected, linguistically different, and culturally distinct.

Corroborating these results on cultural frictions in assessing firm-relevant information and their impact on long-run belief formation, we also find evidence that, on average, forecasters are more

[2] This result highlights the importance of persistent firm variables, such as industry classification, for shaping long-run beliefs. It also aligns with the discussed firm life-cycle component of long-run beliefs, as in-sample, cross-firm variation in firm age outweighs within-firm variation and is therefore largely captured by firm fixed effects.

[3] Brokerage house fixed effects, perhaps surprisingly, contribute negligibly to the explained variation for either forecasting horizon, implying that organizational norms and protocols have a limited influence on analyst expectations.

[4] Additional prior studies also report low explanatory power of demographics for expectation variation, often focusing on more homogeneous samples, such as households from a single country (e.g., Malmendier and Nagel, 2011; Bailey et al., 2019; Ben-David et al., 2018; Kuchler and Zafar, 2019; Das et al., 2020; D'Acunto et al., 2023).

optimistic about firms located in countries they perceive to be more trustworthy, as proxied by European cross-country survey data on trustworthiness. This finding, which is consistent with the notion that trust fosters optimism by shaping expectations about others' competence (Jones, 1996), is robust within firm-year, effectively comparing the TGR beliefs of two forecasters from countries with different levels of trust toward the same focal firm's country. Overall, the evidence in this part of the paper demonstrates that forecasters' cross-sectional characteristics shape differences in long-run growth expectations.

Finally, related to (ii), we link analysts' TGR expectations, and heterogeneity therein, to differences in the thematic focus in the text portions of the equity reports. This provides a rare setting where we observe both quantitative expectations and qualitative discussions of these estimates, allowing us to open the "black box" behind the belief formation process. Specifically, we estimate a latent Dirichlet allocation (LDA) model, and investigate how forecasters allocate topic focus along both the extensive and intensive margins.[5]

This final qualitative-based analysis yields two key insights. The first is that heterogeneity in quantitative TGR expectations is primarily driven by differences in how forecasters prioritize overlapping topics in their discussions, rather than by the selection of entirely distinct topics. We establish this greater importance of intensive over extensive margin topic differences through pairwise comparisons of forecasters who evaluate the same firm simultaneously. This approach allows us to collapse the high dimensional text data space into a single easily comparable measure for each forecaster pair, providing a new angle to explore the role of qualitative information in shaping expectations. The second insight emerges from connecting the qualitative topic focus results back to our extrapolation evidence. Greater geographic distance between forecasters is associated with greater topic focus disagreement on the intensive margin. Additionally, variation in TGR beliefs is particularly driven by differences in topic intensity related to geographic and macroeconomic themes. These findings suggest that local factors shape the degree of emphasis forecasters place on specific topic, serving as a key mechanism underlying extrapolation.

Overall, we provide the first comprehensive analysis of the origins and drivers of (professional) forecasters' very long-run cash flow expectations. Our findings help inform macro-finance models by highlighting the dominant role of individual heterogeneity in explaining long-run—but not short-run—beliefs, i.e., in more abstract contexts malleable to subjective perspectives. A related key model ingredient for long-run beliefs is extrapolation from one's own context, which, as our heterogeneity and topic analyses reveal, operates through differences in when and how much

[5] The LDA approach is preferred for our purposes, as we are primarily interested in identifying thematic structures and variations in analysts' qualitative discussions. In related work, Bastianello, Decaire, and Guenzel (2025) opt for using large language models (LLMs) to extract specific reasonings, narratives, and underlying mental models from forecasters' qualitative discussions, enabling them to link causal relationships in the text to subjective expectations—specifically, in their context, price target expectations.

forecasters attend to local information and signals.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 provides institutional background and introduces the data as well as methodology. Sections 4 and 5 present the results. Section 6 concludes.

2 Literature Review and Contributions

We make four main contributions to the literature. First, we add to the literature on investor beliefs by providing evidence on the (cross-sectional) determinants of truly long-run firm growth expectations, extending beyond near-term horizons of a few years. Research on long-run beliefs is lacking, despite the fact that expectations at long horizons are crucial for asset pricing and valuation. Mukhlynina and Nyborg (2018) and Decaire and Graham (2024) show that forecasters expect over 70% of discounted cash flow to be realized beyond a 5-10 year horizon. Relatedly, Nagel and Xu (2022) and Bordalo et al. (2024b) find that long-term growth expectations, rather than short-term growth expectations, explain much of the stock market valuation cycles.[6] However, these studies rely on IBES LTG expectations (i.e., 3–5 year forecasts), consistent with a broader literature in behavioral finance and asset pricing that uses LTG as a proxy for the long run, going back to at least La Porta (1996). This paper uniquely contributes by providing a detailed analysis of market participants' beliefs—along with their heterogeneity—about firm growth at much longer, truly long-run horizons.

Second, we contribute to the rapidly growing literature on investor and analyst beliefs that draws its foundation from large datasets extracted from analyst equity reports. Decaire and Graham (2024) present numerous stylized patterns about how professionals perform valuation, and provide a deep analysis of the underlying factors shaping how their estimate discount rates in practice. By contrast, we examine the heterogeneity in long-run expectations across forecasters (and firms), leveraging variation in forecasters' backgrounds from matched LinkedIn data and the qualitative discussions in their reports to explain their quantitative beliefs.[7] Other contemporaneous papers based on analysts' equity reports include Decaire et al. (2023), who examine how financial professionals navigate parameter uncertainty in valuation and its connection to financial market outcomes, rather than analysts' long-run expectations, as well as Bastianello et al. (2025), who extract the mental models underlying analysts' financial forecasts from equity reports (cf. footnote 5).

[6] Nagel (2024) explains the differences in conclusions regarding the relative importance of short- and long-run cash flow expectations (e.g., De La O et al. (2024)) by pointing to variations in the definition of earnings used and the degree of noise introduced by the inclusion or exclusion of special items.

[7] By analyzing how truly long-run, TGR expectations are shaped by variation in qualitative discussions across forecasters, we also contribute to the burgeoning literature in economics and finance that integrates quantitative and qualitative information—such as text data underlying quantitative forecasts—to shed light on economic agents' belief formation process (e.g., Kendall and Oprea (2023); Demgensky and Fritsche (2023); Bybee et al. (2023b); Andre et al. (2024); Charles and Kendall (2024); Charles and Sui (2024)).

Third, we contribute to the ongoing debate on the extent to which cash flow expectations about firms may be influenced by current prices. Belief reverse-engineering from prices could result in the misattribution of price variation to beliefs rather than to time-varying expected returns, as summarized in Nagel (2024). For example, in the context of IBES LTG expectations, Chaudhry (2025) argues that instrumented non-news fluctuations in stock prices affect analyst expectations, conduct various tests to refute the notion of spurious correlations between LTG beliefs and prices, including adding controls for expected return proxies and price ratios. Our contribution lies in the unique advantage of our data, which allows us to observe all components of forecasters' belief formation models. This enables us to derive and test novel predictions of the reverse-engineering hypothesis—specifically, that subjective expected returns must exhibit offsetting errors with TGR beliefs to balance the price equation. Empirically, we find no support for this prediction, consistent with TGR expectations reflecting true subjective beliefs rather than being contaminated by prices.

Finally, we contribute to the literature on the macro-finance belief formation process of investors and financial market participants. In recent years, a robust body of work has successfully used surveys to measure subjective expectations of various economic agents across different settings, including beliefs about inflation and the stock market (see, e.g., Coibion and Gorodnichenko (2012), Giglio et al. (2021), Weber et al. (2022), and D'Acunto and Weber (2024)). We complement this literature by incorporating self-reported belief data—along with textual explanations of those beliefs—from forecaster-generated documents (equity reports), rather than researcher-elicited surveys. By documenting the importance of individual heterogeneity and heuristic-driven belief formation in forecasters' very long-run firm growth expectations, we also contribute to prior work on the effect of personal characteristics, experiences, extrapolation, and trust on beliefs, learning, as well as economic outlooks and decisions (e.g., Malmendier and Nagel, 2011, 2016; Bailey et al., 2019; Ben-David et al., 2018; Kuchler and Zafar, 2019; Hartzmark et al., 2021; D'Acunto et al., 2021; D'Acunto et al., 2022a,b, 2023). Relatedly, by focusing on biases in forecasts by individual analysts, we also contribute to the literature in behavioral corporate finance, particularly the less-developed strand focusing on "third parties,"—i.e., non-investors and non-managers (see Malmendier (2018) and Guenzel and Malmendier (2020) for recent surveys of the field).

3 Institutional Details, Data, and Methodology

3.1 Institutional Details

Since we assemble the long-run cash flow expectations dataset from equity reports published by professional analysts (i.e., from the original documents), it is useful to first provide some background on the analytical practices in the equity research industry. Among the methods used to perform valuation, financial textbooks (Berk and DeMarzo, 2019; Brealey et al., 2022) and business

school curricula emphasize discounted cash flow models (DCF) and valuation multiples methods. Echoing these recommendations, these two approaches are the most common tools used by financial professionals to perform valuation (Graham and Harvey, 2001; Mukhlynina and Nyborg, 2018), with one or both strategies included in close to 100% of equity reports housed by our data provider, Refinitiv (see Section 3.2.1). The use of DCF models in equity reports has steadily increased over the past 20 years, accounting for approximately 40% of all reports including DCF analysis over the past 15 years (Decaire and Graham, 2024).

Although the two valuation methods share many similarities, our data collection focuses on reports produced with DCFs specifically for one key reason. In contrast to valuation multiples that only disclose a single metric—the multiple—DCFs require forecasters to report detailed modeling assumptions, allowing us to clearly distinguish the terminal growth rate beliefs of analysts from the other parameters of the valuation model. Appendix Figure B.2 presents an example of a complete DCF model from an equity report, with the numbers used in the model redacted for copyright reasons. This granularity of the data guarantees that we never have to estimate or approximate any of the terminal growth rate expectations. Instead, our analysis is conducted entirely with data directly observed in the analyst reports. This contrasts with prior literature that relies on commercial databases such as IBES or Value Line, which provide only shorter-horizon expectations, forcing previous studies to use imprecise proxies for truly long-run beliefs in their analyses.[8]

3.2 Data

3.2.1 TGR Beliefs Data from Analyst Equity Reports

To construct the long-run cash flow expectations dataset, we initially downloaded 157,549 equity reports mentioning "DCF" from 55 major equity research firms via Refinitiv Eikon. We restricted the time window to reports published in the first quarter of each calendar year (January 1 to April 1) from 2000 to 2023, to ensure that belief data and other variables were systematically measured at consistent time each year. When analysts published multiple reports on the same firm within the first quarter of a given year, we retained only the earliest publication to avoid duplicates within an analyst-firm-year pair. This procedure resulted in 41,429 reports that (1) contained the forecaster's terminal growth rate expectation and, crucially for our analysis of heterogeneity in long-run expectations across forecasters, (2) allowed us to unambiguously identify both the *analyst's name* and *location* at the time of report publication, using landline country indicators for the latter.

[8] As we discuss in detail in Section 4.1, the "long-term" growth estimates reported in these databases (e.g., LTG in IBES) do not reflect truly long-term expectations provided by analysts, as they are an order of magnitude larger than actual TGR expectations. Additionally, we note that, of course, analystsdo not randomly choose between a multiples- and DCF-based method, though we view any resulting selection or external validity concerns as limited in this context. Prior work reports no systematic link between valuation methods and market reactions or price target accuracy but finds evidence that analysts favor DCFs for harder-to-value firms (Asquith et al., 2005; Hashim and Strong, 2018).

We collected the terminal growth rate included in each of these reports in four steps. First, we pre-processed each document using a series of Python programs, including pdf extraction functions and OCR packages, to identify all sections of text, tables, and figures containing relevant information of interest for our purposes. Second, we converted these different elements into short text snippets. Third, for each variable collected, we used artificial intelligence, Open AI API, to extract the numerical value from these snippets. We conducted steps one to three on the Arizona State University High Performance Computing servers *Sol*, as these steps are computationally intensive and require a large number of processing nodes combined with significant storage capacity. Fourth, we exported to Excel the text snippets and the numerical values extracted by artificial intelligence, and manually verified the validity of each extracted value with the help of our research team. This last step of the data collection process is crucial for ensuring the integrity and high quality of the dataset. While artificial intelligence has been shown to be an efficient tool for text extraction Gilardi et al. (2023), and processes complex sentences with a high success rate, one may remain worried about error rates being above acceptable levels when left unsupervised.

3.2.2 Firms and Coverage

The final sample of 41,429 equity reports, coming from 41 of the largest brokerage houses operating throughout the world, covers 8,010 firms located in 49 countries, and include analysts located in 44 countries during the 2000-2023 period (cf. the summary statistics in Table 1, which we discuss in more detail in Section 3.4). The average forecasted firm owns assets with a book value of $15.8 billion USD and has an investment rate of 5.55%. These magnitudes are comparable to the IBES universe, where the average firm has assets with a book value of $15.2 billion USD, and an investment rate of 5.3%. Overall, this suggests that the firms in our sample are of comparable size and invest with similar intensity as those included in other broadly distributed commercial analyst datasets.

The average firm is included in the sample for 7.9 years, and is followed by 7.2 analysts producing DCF reports. In terms of geographic coverage, 39% of firms have their headquarters located in Europe, 27% in North America (24% in the US), 19% in Asia, 12% in Oceania, 2% in South America, and 1% in Africa.

37% of the reports are produced by analysts based in Europe, 30% in North America (27% from the US), 18% in Asia, 12% in Oceania, 2% in South America, and 1% in Africa.

Our sample includes two dozen NAICS industry sectors (2-digit), with the eight largest broad sectors accounting for 84% of the total coverage: 35% for manufacturing (NAICS 31-32-33), 19% for information (NAICS 51), 8% for professional services (NAICS 54), 7% for retail trades (NAICS 44-45), 4% for mining and oil & gas (NAICS 12), 4% for transportation (NAICS 48-49), 5% for utilities (NAICS 22), and 4% for finance and insurance (NAICS 52).

Overall, these statistics suggest that our sample is comprehensive and broad, representative, and comparable to its commercial counterparts.

3.2.3 Analyst Backgrounds from Universe of LinkedIn Data

We merge the TGR dataset on long-run expectations described in Section 3.2.1 with the universe of LinkedIn users obtained from Revelio Labs, thereby obtaining additional (self-reported) personal characteristics of forecasters, including their educational background and (approximate) age. The merge is made possible since as described, the TGR dataset contains key identifiable information on the forecasting analysts, including their full name and the name of their employer, allowing for a clean and unambiguous merge with the Revelio data.

In linking the datasets, we require the names across individuals in the TGR and LinkedIn datasets to match precisely. To further err on the side of caution in constructing the matched dataset, we also require the LinkedIn-provided employment history to match the analyst's employer in the equity reports,[9] as well as the name-employer matching to result in a unique LinkedIn user that is matched to the data.[10] This procedure leads to 1,667 uniquely matched analysts across the TGR and LinkedIn datasets (Panel A of Table 1) out of the 5,300 equity analysts included in our sample (31%).

3.3 Methodology: Latent Dirichlet Allocation Estimation

To identify the topics discussed in reports, and quantity forecasters' attention allocation across topics, we employ a Latent Dirichlet Allocation (LDA) model. The LDA model estimation is performed in two steps: (i) text pre-processing, and (ii) statistical estimation.

3.3.1 Text Pre-Processing

We start by extracting the text from each report using a series of Python routines. Specifically, in this step, we *exclude* information contained in figures, tables, and the legal boilerplate disclaimers at the end of reports. The objective is instead to capture the valuation thesis that professionals articulate to support their numerical analysis. In a final step, we do a final pass on the extracted text to address extraction issues that might have been missed by the Python routine by using Meta's LLM, Llama-2-70b.[11]

Then, we apply five transformations to the text data. First, for an equity report to be included in the LDA estimation, it must contain at least 100 words. Second, we remove verbs from the text.

[9] We account for various commonly used employer names (e.g., Bank of America or BofA) as well as accented characters (e.g., Societe Generale or Société Générale).

[10] That is, we drop "matches" of multiple LinkedIn users with the same name and who list the same employer in their LinkedIn biography.

[11] For example, in some cases, the extracted text does not contain spaces between words. LLMs are efficient at breaking sentences into individual words.

Third, we remove stopwords and lemmatize words using standard Python libraries. Fourth, we organize words into unigrams and bigrams, based on the terms' occurences in our full set of reports. Pairs of words that often arise together receive higher likelihood of being treated as bigrams than as two unigrams.[12] Last, for a unigram or bigram to be included in our final dataset, the term has to appear at least once in 0.1% of all the reports used in the analysis.

3.3.2 LDA Model Estimation

Our estimation strategy is similar in spirit to Bybee et al. (2023a)'s approach, to which we refer readers interested in the estimation technicalities. At an intuitive level, the LDA model uses a bag-of-words approach where each report can be represented in an article-term matrix of dimension $R \times T$. Row indices correspond to the list of R different equity reports, and T is the column indices that correspond to the distinct terms that represent the corpus generated from the text pre-processing steps. Numerical values of individual elements in the article-term matrix then correspond to the number of occurrences for each term in each report t.

To estimate an LDA model, one must select three hyper-parameters: (i) the number of topics (K), a parameter governing the distribution of topics within reports (α), and (iii) a parameter determining the distribution of terms within a topic (η). We set α to be equal to 0.05, and η to 0.02. These values were obtained when optimizing the perplexity score of our selected model. These magnitudes are also broadly consistent with the estimation standard for LDA models, where a simple rule of thumb such as $\frac{1}{\text{Number of topics}}$ is generally a reasonable starting point for the estimation.[13]

To determine the optimal number of topics (K), we estimate different versions of the LDA model using from 10 to 200 topics, in increments of 10 topics. Our model selection strategy for the number of topics is based on three metrics: (i) the coherence score, (ii) the perplexity score of the LDA model, and (iii) a manual investigation of the keywords contained in each topic. The coherence score provides a measure of the topic model's interpretability by comparing the semantic similarity of the words included in each topic. A higher coherence score indicates greater semantic similarity across the words in each topic. The perplexity score indicates how well the model describes our set of documents, with a lower perplexity score suggesting a better fit of the data. While these metrics are useful in guiding model selection, we also manually examine the set of words included in each topic to confirm their *natural* interpretability.

Figure B.1 plots the coherence and perplexity scores over the range of possible numbers of topics included in the LDA (Appendix Table B.2 reports the keywords included in our favored model.) As the figure shows, most of the improvements in coherence and the in-sample perplexity scores are

[12] For example, when "supply" and "chain" are written consecutively in a text, the likelihood of using the bigram "supply chain," instead of their unigrams is higher.

[13] We note, however, that the conclusions of the tests using this output are not sensitive to perturbations around those magnitudes.

obtained when reaching 50 topics, and after that improvements level off. Upon manual inspection, we also find that the 50-topic model is parsimonious, and contains topics that are meaningful, mostly non-overlapping, and distinctive.

We also note that the coherence score is maximized at 150 topics. However, upon manual inspection of the set of words used to describe each topic, we find them to be more generic, with some overlapping topics. Nonetheless, while our favored LDA specification uses 50 topics to model reports, we use the 150-topic model to validate the robustness of our main analysis.

3.4 Summary Statistics

Table 1 presents detailed summary statistics. As alluded to above, our TGR dataset contains more than 40,000 TGR forecasts by over 5,300 analysts that make forecasts for approximately 8,000 unique firms (Panel A).

In Panel B, the average TGR forecast in the data is 2.2%. Panel A of Figure 1 plots the average TGR across firms and analysts over time, showing that average long-run expectations have somewhat declined over time, from around 2.8% at the beginning of the sample in 2000 to closer to 2.0% towards the end of the sample. In terms of levels, the TGR expectations are an order of magnitude smaller than the average IBES-based LTG expectation for the sample firms of 13.0%. (We will contrast TGR and LTG expectations further in Section 4.1.) Also shown in Panel B of Table 1, about 1 in 3 equity reports and TGR forecasts come from foreign analysts, i.e., analysts who forecast the long-run growth of a firm headquartered in a different country than where the analyst is located.

Panel C contains demographic and other information about forecasters. Just under 20% of analysts are female. The vast majority of analysts are Caucasian (75%) as per the Revelio-based race prediction algorithm. The next largest group is Asian analysts (18%). 44% of analysts have a graduate degree (Master's and/or MBA). There is considerable variation in analyst age and birth years, with the 25th-percentile analyst born in the early 1970s and the 75th-percentile analyst born in the mid-1980s. In terms of other analyst-related statistics, the average analyst in the sample contributes 8 reports and TGR forecasts. There are about 120 analysts on average who make forecasts for firms located in a given country, and 140 analysts per brokerage house.

Panels D and E present descriptive statistics on various economic series for the countries in which forecasted firms and analysts (conditioning on foreign-based analysts) are located. There is meaningful variation in economic conditions of both firm and analyst countries. For example, the interquartile range (IQR) in firm-country real GDP and inflation is 2.2pp and 1.9pp, respectively. Similarly, the IQR in analyst-country GDP and inflation is 1.7pp and 1.4pp, respectively.

4 The Relevance of TGR Expectations

This section examines the economic importance of analysts' long-term cash flow expectations through four key steps. First, we demonstrate that terminal growth rate (TGR) expectations provide distinct economic insights beyond those captured by other expectation series used in prior research. Specifically, we show that TGR expectations exhibit limited overlap with the traditional long-term expectation benchmark, IBES LTG. While TGR expectations are strongly associated with ex-post realized long-term firm earnings growth, IBES LTG—designed to capture analysts' expectations over a 3- to 5-year horizon—better captures medium-term dynamics. Second, we establish that the cross-sectional determinants of TGR expectations differ from those driving short-term forecasts. Third, we show that TGR expectations are anchored in theoretical determinants of long-run growth. Finally, we present novel evidence indicating that TGR expectations are not contaminated by investors' expected returns.

4.1 Properties of TGR and IBES LTG Expectations

We start by examining the relation between terminal growth rates and the IBES LTG variable, and establish that both variables share limited similarities. Panel B of Table 1 reports that the correlation between the two series is weak ($\rho = 0.07$). Contrasting the time trends in both variables, we find that the terminal growth rate yearly average has steadily declined over the sample period (Panel A of Figure 1), whereas the average level of IBES LTG has remained constant. Further, comparing the levels of both series (Panel A of Figure 2), we observe a systematic and persistent gap between the two series, averaging 10.0 pp throughout the sample period, and showing that TGR expectations are substantially smaller than IBES LTG expectations. Lastly, Panel B of Figure 2 shows a scatter plot representing the relation between IBES LTG and terminal growth rates. Consistent with the weak correlation presented in Panel B of Table 1, it is difficult to visually discern a clear relation between both variables, given their considerable degree of dispersion. The figure also clearly shows that IBES LTG exhibits substantially more volatility than the terminal growth rate. Table 1 confirms the visual inspection of the scatter plot and shows that the standard deviation of IBES LTG (22.5 pp) is 17 times greater than that of the terminal growth rate (1.29 pp).

Adding to the differences between the TGR and LTG expectation series, in Appendix Figure B.3, we also find that forecasters' long-term TGR stock growth expectations are systematically below historically realized long-term growth rates in every year of our sample, whereas LTG expectations studied in prior work are systematically above historical growth rates (Nagel and Xu, 2022; Bordalo et al., 2024b).

Altogether, the evidence in this section establishes that TGR expectations are systematically different from other expectation series that target growth beyond a one- or two-year horizon, such as

the IBES LTG series. This implies that our existing understanding of how professional forecasters and market participants form long-term expectations about firms' growth potential remains limited.

4.2 TGR and Future Realized Growth

To directly establish TGR as a meaningful measure of long-run growth expectations, and the distinguish differences in the economic information content of TGR and LTG expectations, we study how these measures relate to firms' future realized long-term growth at multiple horizons.

We first reiterate that terminal growth rates are inputs from analysts' DCF models. Decaire and Graham (2024) shows that the modal DCF model sets the explicit forecast horizon to be equal to three years and then applies a constant terminal growth rate to any subsequent cash flows. Against this background, in our main specification, our goal is to capture growth expectations from period $t+3$ and onward, when the modal analyst starts applying the terminal growth rate in the DCF model. Our main measure of realized growth is then the geometric average of the ex-post realized earnings growth from year $t+3$ to $t+8$, measured for each firm in our sample. Comparing the TGR beliefs with this ex-post measure of realized growth in Panel C of Figure 1, we find that both series strongly comove together, but note differences in their respective levels (i.e., TGR expectations − ex-post realized growth = −3.2 pp). The fact that both series share a similar trend is reassuring, but the difference in levels calls for additional tests that we conduct at the end of this section.

To formalize the pattern documented in Panel C of Figure 1 result, we run a linear regression where we consider three possible alternative ways to measure long-run ex-post realized growth: the average geometric earnings growth for each firm in our sample over the periods $t+1$ to $t+6$, $t+3$ to $t+8$, and $t+3$ to $t+13$. The last window expands the horizon over which we measure realized growth to account for a 10-year period, at the expense of a reduced sample size. Panel A of Table 2 presents the results.

Focusing on our preferred specification in Column (2), which measures realized growth over the $t+3$ to $t+8$ horizon, we find that a one percentage point increase in TGR expectations is associated with a 0.66 percentage point increase in ex-post realized growth. This strong, positive relationship between TGR beliefs and realized growth persists across other examined horizons, as shown in Columns (1) and (3). The robustness and persistence of this relationship support the interpretation that, for the modal model, TGR reflects the average realization of earnings growth from $t+3$ onward, effectively capturing the firm's expected long-term growth trajectory.

This interpretation is consistent with observed data patterns: realized growth rates stabilize as firms mature (Figure B.4). Moreover, analysis of annual average growth measures across firms, presented in Appendix Table B.4, shows that TGR expectations account for a significant share of ex-post realized growth, with an R^2 of 0.45. Together, these findings highlight the pivotal role of TGR in reflecting firms' long-term growth expectations.

Further, to distinguish between terminal growth rate (TGR) and long-term growth (LTG) economic content, we incorporate the IBES median consensus LTG variable into our analysis. Including this variable allows us to compare how the relationships of TGR and LTG with future realized earnings growth evolve as we extend the measurement horizon for growth.

Turning to the relationship between long-term earnings growth and IBES LTG forecasts, we uncover a markedly different pattern. At the IBES LTG-stated forecast horizon of 3–5 years (Bordalo et al., 2024b), we find a meaningful economic relationship: a one percentage point increase in LTG expectations is associated with a 0.30 p.p. increase in the average realized earnings growth over the $t+1$ to $t+6$ horizon. However, as the horizon for measuring realized growth extends further into the future, the relationship steadily and monotonically weakens. By the $t+3$ to $t+8$ horizon, the association declines to just 0.08 percentage points. This pattern stresses the diminishing predictive power of LTG forecasts as the horizon lengthens, indicating that LTG expectations are increasingly disconnected from realized earnings growth outcomes over longer time frames.

Broadly, the results of this section show that both TGR and LTG are associated with long-run realized earnings growth, but the nature of their relation, especially as we extend the horizon evolves in different ways confirming that both variable information content is focused on different horizon of firm future growth. This disconnect between shorter-run and long-run expectations should not come as a surprise. Indeed, in the data, we document the same relation between realized growth rates (Figure B.5).

5 Drivers of Expectations: Short-run Versus Long-run

Having established (i) that TGR expectations have distinct properties compared to other long-term expectations variables studied in the existing literature, (ii) that it targets growth at longer horizons thatn previously studied expectation series, and (iii) that they predict realized long-run firm growth, we now investigate and compare short-run and TGR key cross-sectional drivers. Specifically, we employ a variance decomposition of TGR beliefs, and link the (heterogeneity in) beliefs to a host of underlying firm as well as analyst characteristics.

5.1 Variance Decomposition

Table 4 decomposes the variation in TGR expectations across brokerage houses, analysts, and forecasted firms, using an ANOVA-based sequential variance decomposition. Panel A estimates the decomposition on the full sample excluding firms, analysts, and brokerage houses for which we have less than 2 observations. Panel B excludes firms with fewer than ten observations for robustness. In both panels, we include specifications with at least 10, 25, and 50 observations per forecaster, to address potential concerns about overfitting.

We consistently find that the three sets of fixed effects explain a large portion of between 60 and slightly above 70 of the variation in TGR expectations. Brokerage house fixed effects contribute only a very small share to the explained variation, whereas analyst and firm fixed effects drive most of the explained variation. For example, in the largest sample in Column (1) of Panel A, brokerage house fixed effects contribute 2 of the explained variation, whereas analyst and firm fixed effects contribute 0.71 and 0.27, respectively. Similarly, in the most narrow sample in Column (4) of Panel B, we estimate contributions of 15, 52, and 33, respectively.

One limitation of the approach in Table 4 is that the ANOVA-based estimation is not systematically indifferent to variations in the sequential inclusion order of the brokerage, analyst, and firm fixed effects. As an alternative approach, we therefore estimate standard OLS regressions in which we gradually include the various fixed effects one at a time in Appendix Table B.6. Panel A and B respectively replicate Table 4 Column 1 of Panel A and Column 4 of Panel B, the two most extreme cases of that table. Columns 1-3 present the results for including each fixed effects individually, effectively estimating an *upper bound* for the amount of variation that can be explained by the inclusion of the given fixed effects. Intuitively, the included fixed effects may be correlated with other (omitted) variables that also affect TGR expectations (cf. the related discussion in Ayyagari et al., 2008).

Using the alternative OLS approach, brokerage fixed effects alone can explain a very small share of at most 2% of the total variation in TGR expectations. Analyst fixed effects alone can explain a much larger share of up to 31% of the TGR variation. Firm fixed effects alone can explain a similarly large share of up to 33% of the variation. It is also worth noting that brokerage and firm fixed effects jointly can explain up to 34% of the variation, whereas including in addition analyst fixed effects leads to an incremental R^2 of 19% (i.e., a total R^2 of 0.53). In a final effort to ensure that our conclusion are not driven by overfitting the data with the fixed effects high degrees of freedom, we look at the adjusted R^2, as a way to penalize the number of degrees of freedom used by each specific fixed effects. Focusing in the adjusted R^2 yields similar conclusion. Specifically, we find that adding analyst fixed effects to brokerage house and firm fixed effect increase the adjusted R^2 by 15%, from 24%. The conclusion we derive from studying the adjusted R^2 in the OLS specification echo the findings we obtain when evaluating the partial adjusted R^2 we estimate in the Anova regression.

We then extend the ANOVA variance decomposition to short-term growth expectations–IBES 1-year EPS growth forecasts–to examine whether the importance of cross-sectional drivers differs between short- and long-run forecasts. The results, presented in Columns 5–8 of Panels A and B in Table 4, reveal significant differences. Forecaster fixed effects explain a substantially larger share of the explained variation in TGRs compared to short-run expectations, systematically contributing 22 to 55 percentage points more. Similar to the results for TGRs, brokerage house fixed effects explain a negligible portion of the variation, while firm fixed effects remain a key factor. These findings

suggest that individual forecaster characteristics play a systematically greater role in explaining long-run expectations than they do for their short-run counterparts.

Overall, the variance decomposition results highlight the importance of firm factors and individual forecaster characteristics in shaping professionals' long-run cash flow expectations. These findings motivate a deeper analysis of both firm characteristics as well as forecaster demographics and backgrounds for explaining the large and persistent heterogeneity in TGR expectations across firms and analysts.

5.2 Firm Characteristics and Theorized Drivers of Long-Run Growth

In this section we explore how TGR expectations relate to firm characteristics. A natural approach is to associate the cross-firm heterogeneity in TGR expectations with factors related to the firm life cycle, as it is well documented that, as firms age and an entire industry becomes more mature, products become more standardized and opportunities for product innovation diminish (e.g., Klepper, 1996).

We first relate TGR expectations to firm age in Panel B of Figure 1, which shows that forecasters' TGR expectations monotonically decline with firm age. The average TGR in the sample is 2.5% for young firms aged 10 years and less, and declines to 0.5% for very old firms, reflecting diminishing growth opportunities with firm age. Importantly, these magnitudes reflect the relation between TGR and firm age cohort *after* controlling for firm and year fixed effects. Columns (3) to (5) of Table 3 Panel B confirm that the decline in TGR expectations with increasing firm age is statistically significant using an OLS regression.

Motivated by this evidence, we then explore when the explanatory power of firm characteristics are most pronounced, and how the importance of those features evolve over firm life-cycle.

Table 5 examines how the variation in TGR expectations that is explained by industry, firm-country, and firm factors net of industry and country fixed effects, respectively, changes across the firm's life cycle. In the left column, we find that a firm's industry and country explain only a small portion of TGR expectations in young firms. Specifically, these factors explain only 7% of the variation in very young firms aged 5 years or less. Even 20-25 years after a firm's IPO, industry and country explain only about 23% of TGR forecasts. In the right column, we find that in these young firms, firm fixed effects explain a large share of the variation in TGR expectations that is unaccounted for by analyst, brokerage house, industry and firm country factors, consistent with the evidence in Section 5.1. They account for up to 37% of unexplained variation in very recently IPOed firms, and up to about 27% for firms aged 20-25 years.

These patterns evident in young firms reverse as firms mature. In old firms, Column 2 of Table 5 shows that a firm's industry and country can explain a much larger amount of the variation in TGR expectations, e.g., 49% for 50-year-old firms and an even larger magnitude for 70+-year-old

firms. Correspondingly, for mature firms, the incremental explanatory power of firm fixed effects over industry and country strongly diminishes.

This evidence highlights that the role and nature of firm-specific factors in explaining TGR heterogeneity evolve systematically over the firm life cycle. We interpret these findings through the lens of a firm life-cycle framework, wherein forecasters grapple with uncertainty about whether a new firm will mature into a Wall Street success or succumb to bankruptcy in the near term.[14] As forecasters gain information about a firm's long-term viability, expectations for average long-run growth increasingly align with broader country and industry trends.

As an example of firm-specific factors that is correlated with TGR, Panel B of Table 3 presents evidence about the role long-term default risk. To evaluate how different levels of default risk impact the terminal growth rate for firms operating in the same country and industry during the same year, our specification includes firm-country*industry*year fixed effects. We also control for firm age, measured as the natural logarithm of the firm's post-IPO age, since our evidence shows that firm age is an important drivers of TGR expectations that likely correlates with default probability. Using this specification, we estimate a negative link between TGR expectations and default risk, indicating that analysts appear to apply a 0.21 pp discount, a 10% reduction compared to the sample average, to firms that have a probability of default above the sample median. Considering that, conditional on being in the *high* risk of default group (above the median), the average probability of default is equal to 5.8%, the 10% reduction in the terminal growth rate is commensurate.

Interpreting the downward trend in TGR values documented in Panel B of Figure 1 within this life-cycle framework suggests that forecasters' optimism about the growth potential of young firms outweighs potential negative revisions. Such pattern aligns with the general over-optimism in longer-run forecasts documented in Bordalo et al. (2024b).

While we report the magnitudes for very old firms (aged 55 years and older), we caution that the sample size for these cohorts is relatively small—fewer than 100 observations—raising concerns about overfitting risks. Despite this limitation, the patterns observed for older firms are consistent with the progression we observe among the younger cohorts, where sample sizes average several thousand observations per age group. Specifically, as firms age, country- and industry-level factors play an increasingly dominant role, while the influence of firm-specific factors, conditional on other observables, diminishes to negligible levels.

The evidence indicates that the source of firm fixed effects explanatory power evolves over the firm life cycle (see Section 5.1). Specifically, idiosyncratic firm-specific factors account for a larger share of the variation among younger firms, while broader characteristics, such as country and

[14] Nearly 25% of public firms ultimately declare bankruptcy, with most failures occurring within 9.4 years of their IPO. This statistic is derived from North American Compustat data spanning 1950 to August 2023 and includes firms classified as bankrupt (code 02), liquidated (code 03), or active, excluding others based on the mnemonic DLRSN.

industry factors, become increasingly dominant as firms mature.

5.2.1 Dispersion in TGR Expectations Over Firms' Life Cycle

As a bridging step between studying firm and analyst characteristics, we also examine how the within-industry-year *dispersion* of TGR expectations across analysts varies over the firm's life cycle. As Table 6 shows, dispersion in analyst TGR beliefs is largest in young firms. For these, the estimates imply that the average absolute dispersion in TGR expectations in the sample is about 1 pp. The TGR belief dispersion across forecasters declines markedly as firms mature, by as much as 85% for the most mature firms.

These patterns are robust to specifications with and without industry-by-year fixed effects as well as a control variable for firm size, and we note that a Bonferroni adjusted test for the joint equality of all the firm age coefficients is rejected at the 1% level in all specifications. These patterns are consistent with analysts exhibiting greater variation in their access to, processing, and interpretation of information related to firm factors compared to industry and country factors that impact their long-term expectations.

5.3 Reverse-Engineering and Returns Contamination

In this section, we examine whether forecasters derive long-term growth expectations by reverse-engineering them from current market prices. Reverse-engineering, in itself, is not necessarily problematic if forecasters accurately extract investors' true long-term beliefs from prices. In such case it would be equivalent to forecasters independently forming long-term expectations while employing a mental model similar to that of investors. In both scenarios, TGR would convey meaningful insights about the behavior of key financial agents.

The underlying concern with the reverse-engineering critic is that the econometrician does not know whether investors and forecasters share similar valuation models, and their associated inputs. Precisely, if forecasters discount rates differ from the one used by investors, this could lead to a TGR measure that is contaminated with investor returns. In this case, the negative correlation documented in the literature (e.g., Bordalo et al. (2024b)) between long-term expectations and future realized returns would not be due to investors and forecasters over-optimism, but rather simply due to forecasters improperly extracting investor long-term beliefs from market prices. Namely, forecasters would incorrectly attribute high prices today to investors optimistic long-term beliefs, whereas high prices today would be truly driven by low discount rates (i.e., investor expected returns).

First, to establish a basis for this discussion, we show that TGR is negatively related to future realized stock returns. Similar to Bordalo et al. (2024b), who document a negative correlation between medium-term growth expectations and future returns, we find that long-term growth

expectations are similarly negatively correlated with future realized returns at the firm level (Panel A of Table 2). This relationship persists even after accounting for year, country-year interactions, and firm fixed effects. Year and country-year fixed effects control for common shocks affecting all firms and region-specific shocks, respectively. Firm fixed effects address concerns that discount rate measures may lack time variation for certain firms, allowing us to evaluate whether innovations in discount rates align with innovations in future realized returns. These findings extend and replicate the results of Bordalo et al. (2024b) by incorporating a long-term growth variable, IBES LTG, while also building on the predictive regressions between subjective cost of equity and realized returns reported in Decaire and Graham (2024).

Second, we review evidence suggesting that investors and forecasters rely on similar frameworks when forming beliefs about asset prices. A large set of evidence points to the fact that investors likely attribute price variation at least in part to changes in expected returns and earnings growth, and that dividends, earnings, or cash flows help predict returns when using market data (Golez, 2014; Sabbatucci, 2022; Hillenbrand and McCarthy, 2024; Pruitt, 2024). Similarly, using analysts subjective data, recent work indicates that both subjective expected returns and expected earnings growth both play a role in explaining market prices (Chen et al., 2013; De La O and Myers, 2021; Decaire and Graham, 2024). Closely related to our context, Decaire and Graham (2024) finds that forecasters reporting target growth rates (TGRs) structure their models using discounted cash flow (DCF) frameworks and employ time-varying discount rates. Similarly, numerous studies document that the capital asset pricing model (CAPM) is among the most commonly used tools for estimating expected equity returns by both investors (Guercio and Tkac, 2002; Barber et al., 2016; Berk and van Binsbergen, 2016) and forecasters (Brav et al., 2005; Andrei et al., 2023; Decaire and Graham, 2024). Recent evidence also suggests that analysts' discount rate estimates reasonably capture the average realized future returns of stocks (Decaire and Graham, 2024), suggesting that it capture inherent features associated with stock market participant expectations. Taken together, these findings indicate that while some differences may exist in how investors and forecasters evaluate stocks, their underlying conceptual frameworks and valuation tools are likely to overlap meaningfully.

Using this combined evidence, we can formalize the valuation efforts of forecasters in our sample as follow:

$$p_t - d_t = E^* \sum_{j=1}^{N} \rho^j \Delta d_{t+1+j} + \frac{\rho^N}{\rho} E^* TGR - \frac{1}{\rho} r_t^a \tag{1}$$

Here, r^a is the analysts' subjective discount rate, which is time-varying. $E^* \sum_{j=1}^{N} \rho^j \Delta d_{t+1+j}$ is the analyst subjective discrete earnings growth forecast over N periods, followed by a constant growth expectation, E^* TGR, from time $N+1$ to infinity.

Working from Equations (1), the concern raised by the reverse-engineering critic is that forecasters do not use a similar *version* of investors' asset pricing model, such that:

$$p_t - d_t = \frac{1}{\rho} r_t^a + \epsilon + [E^* \sum_{j=1}^{N} \rho^j \Delta d_{t+1+j} - \phi\epsilon] + \left[\frac{\rho^N}{\rho} E^* TGR - (1-\phi)\epsilon\right] \tag{2}$$

Where ϵ is the difference between analysts and investors discount rate, and this *difference* then contaminates their growth expectations as they reverse-engineer investors' growth expectations from observed prices. Conceptually, the contamination from ϵ in growth expectations can be distributed across short- and long-horizon growth expectations so that Equation (2) holds during the reverse-engineering exercise.

Two scenarios are possible: (1) all the contamination from analysts using a different discount rate is contained in short-term growth expectations ($\phi = 1$), or (2) the contamination affects long-term expectations as well ($\phi \neq 1$). Case (1) is not an issue in our analysis, but case (2) would be problematic.

Examining Case (2) closer, ϵ also predicts a direct mechanical relationship between analysts' discount rate and their TGRs. This implies that, even after controlling for macroeconomic factors likely to affect both r_t^a and TGR (e.g., GDP, inflation, and interest rates), there should still be a mechanical link between the two variables in the time series driven by ϵ—which leads to an empirically testable hypothesis: the absence of comovements would provide evidence against a meaningful role for ϵ, after controlling for macroeconomic factors likely to jointly drive variations in both variables.

We provide evidence against a meaningful role for ϵ in Table 2, Panel B. We start by replicating the results presented in Decaire and Graham (2024) in Columns (1) and (2), but using the subjective cost of equity instead of the subjective discount rate. Column (1) investigates this relationship without any fixed effects and finds a small, positive, and statistically significant coefficient, similar to the one documented in Decaire and Graham (2024). This is consistent with the notion that both TGR and the subjective cost of equity may be associated with some of the same macroeconomic variables (e.g., the interest rate).

Next, Column (2) introduces firm-country*year fixed effects, controlling for macroeconomic factors, and firm fixed effects to examine how both TGR and the subjective cost of equity are updated over time. As in Decaire and Graham (2024), we document a small regression coefficient, but in this case, it is not statistically significant for the subjective cost of equity.

Finally, in our most restrictive specification, we include analyst-firm fixed effects, which allows us to evaluate how a given analyst updates both inputs for a particular firm over time. We obtain a

small coefficient that is not statistically significant at conventional levels (p-value > 0.10).[15]

Considering that the relation implied by Equation (4) predicts a direct mechanical—and thus statistically strong—relationship between the two variables, this evidence is potentially consistent with two non-mutually exclusive conclusion: (i) forecaster engage in reverse-engineering, but the long-term expectations they extract are not contaminated by investor expected returns, or (ii) reverse engineering long-term growth expectations from market prices is not widespread among the analysts, and that its influence is negligible. Under both scenarios, this test complements and extends Bordalo et al.' (2024b) efforts showing that forecasters longer-term expectations are unlikely to be contaminated by investors expected returns, and that they contain meaningful information.

5.4 Forecaster Characteristics and Backgrounds

Having established the importance of forecaster fixed effects in Section 5.1, this second part of the analysis examines to what extent analyst observables can explain the persistent individual heterogeneity, making use of the fact that in our data we observe forecaster identities and LinkedIn-matched background information.

Table 7 performs a sequential ANOVA decomposition of estimates of individual analyst fixed effects, which we obtain using a mover design—i.e., analysts moving brokerage houses in our sample, allowing us to separate analyst and employer fixed effects—following the literature in labor economics (Abowd et al., 1999). Jointly, analyst demographics can explain about 30% of the variation in analyst fixed effects, is a larger amount relative to prior studies on retail investors (cf. Giglio et al., 2021) and households (cf. Malmendier and Nagel, 2011; Bailey et al., 2019; Ben-David et al., 2018; Kuchler and Zafar, 2019; Das et al., 2020; D'Acunto et al., 2023), which typically finds that demographics explain at most 10% of the belief variation.

Several factors may explain the greater capacity of demographics to explain heterogeneity in forecaster expectations in our setting. First, our panel data includes multiple observations per forecaster, with most forecasters tracking multiple firms over several years, enabling us to obtain more precise estimates of forecaster fixed effects from of the firm and brokerage house patterns. Second, our sample is more diverse, comprising analysts from multiple countries. This diversity allows us to disentangle the effects of firm and analyst country characteristics, unlike prior studies that often rely on more homogeneous samples, such as affluent U.S. retail investors affiliated with Vanguard (Giglio et al., 2021). Finally, the stickiness of long-term expectations, which exhibit limited time-series variation and are largely cross-sectional (Decaire and Graham, 2024), likely amplifies the role of persistent individual characteristics compared to the more volatile nature of

[15] The conclusions derived from the regressions in Table 2, Panel B hold when TGR and r_t^a are expressed in percentages (i.e., TGR (%) and R_t^a (%)), as shown in the table, or in logarithmic format (i.e., ln(1+TGR) and ln(1+r_t^a)). This robustness rules out the concern that *Analyst error* could be purely multiplicative or additive in our data.

short-term expectations.

Table 7 highlights several demographic factors that are particularly associated with analyst fixed effects in our sample, including race, birth year (age), and location (country of residence at the time of the report). This relationship holds across various specifications: the full sample of analysts with identified individual fixed effects based on Abowd et al. (1999) (Column 1), the subsample of firms with coverage from foreign analysts (Column 2), and the subsample excluding domestic analyst coverage (Column 3). Restricting the sample to firms with some foreign coverage, and excluding domestic coverage helps disentangle the effects of firm and analyst geographic location, ensuring these factors are not confounded.

Lastly, to ensure that the conclusions derived from the variance decomposition are robust to both the sample used in Columns 1–3 and our methodological approach, we conduct two robustness tests. First, Columns 4–6 of Table 7 apply stricter filters to the sample, requiring a minimum of 5 observations for each firm, brokerage house, and analyst to be included in the conservative sample. The results show similar magnitudes, suggesting that our conclusions are unlikely to be driven by a small number of observations per fixed effect.

Second, to examine whether our results depend specifically on the use of an ANOVA-based variance decomposition, we incrementally add fixed effects to OLS regressions to evaluate the corresponding increases in R^2. The results are presented in Appendix Table B.6. Panels A and B replicate, respectively, Columns 1 and 6 of Panel B in Table 7, representing the two most distinct specifications of the analysis. Columns 1–5 in Appendix Table B.6 sequentially introduce each fixed effect individually. Examining the R^2 values, we find that birth cohort (R^2 = 0.09–0.13) and analyst geographic location (R^2 = 0.07–0.13) substantially explain variation in analyst fixed effects, while other demographic variables have negligible explanatory power. Furthermore, adding analyst geographic location to a model already including all other fixed effects—gender, education, race, and birth cohort—significantly improves the explanatory power, with R^2 increasing by 0.03–0.14 percentage points, representing a 60–93% improvement. This finding suggests that analyst geographic location provides incremental information not captured by the other factors.

To further validate the informational relevance of demographic factors, we analyze the adjusted R^2, which penalizes the inclusion of additional variables by accounting for degrees of freedom, thereby reducing the risk of overstating the explanatory power of the model. Using this adjusted metric, we find that analyst geographic location retains substantial explanatory power for variation in analyst fixed effects, with an adjusted R^2 ranging from 0.03 to 0.05. Importantly, incorporating analyst geographic location into a model with other fixed effects further enhances the adjusted R^2, increasing it by 0.03 to 0.08 percentage points—an improvement of up to 75%.[16]

These findings are consistent with the results based on the partial adjusted R^2 presented in Panel

[16] The percentage improvement cannot be computed for Panel B, as the adjusted R^2 in Column 6 is negative.

B of Table 7, where analyst geographic location emerges as the single most influential variable, even after adjusting its partial R^2 for degrees of freedom. Collectively, this evidence highlight the pivotal role of geographic location in shaping analysts' long-term expectations formation.

5.5 Forecasters' Personal Economic Environment and Extrapolative Expectations

A natural question is why analysts from certain backgrounds—such as in terms of their geographic origins and exposures—have systematically higher or lower TGR expectations. This section explores the effect of locally experienced economic conditions on analysts' long-run expectations, including the formation of expectations for geographically distant firms.

Panel A of Table 8 reports estimates examining how personally and locally experienced economic conditions influence forecasters' long-term expectations about firms. This analysis leverages the granularity of our data on the geographic locations of both firms and the analysts forecasting them. In Columns (1) and (3), we find that recent GDP growth in the analyst's country of residence strongly correlate with their long-term expectations for firms. Importantly, this relationship holds even when we control for the recent economic performance in the firm's country of location. [17] These results suggest that analysts substantially extrapolate from local economic conditions when forming their TGR beliefs.

Notably, the effects of home-country experiences are concentrated in forecasts for firms located in geographically distant regions (Column (3)). This pattern implies that local extrapolation may reflect analysts' responses to higher information-processing costs and limited familiarity with firms in distant locations. Such behavior is inconsistent with alternative explanations involving common shocks across neighboring countries. To further address potential confounding mechanisms, we include analyst-country-by-firm-country fixed effects, which control for specific linkages between analysts and firms. For instance, if analysts are systematically more likely to forecast firms with operations in their home country, these fixed effects ensure that the role of local extrapolation is evaluated within such dynamics. The addition of these fixed effects strengthens the interpretation of our findings as evidence of local extrapolation influencing analysts' beliefs.

To refine our understanding of why analysts engage in local extrapolation for distant firms, we examine three dimensions of familiarity and information access and processing cost toward foreign countries: (i) language similarity, (ii) economic ties, and (iii) cultural affinity. We find that local extrapolation is more pronounced when analysts face greater barriers in these dimensions. Specifically, extrapolation is concentrated among analysts who do not share a common language with the firm's country, operate in regions with weak economic ties to the firm's location, or have

[17] Columns (2) and (4) reveal that this extrapolation mechanism operates through macroeconomic productivity factors rather than inflation: we find no evidence of local inflation influencing long-term expectations. This finding aligns with prior evidence in Decaire and Graham (2024), which shows no relationship between TGR and inflation.

limited cultural exposure to the firm's environment.

Language Similarity. Columns 1 and 2 of Panel B in Table 8 split the sample based on whether the dominant language of the firm's country matches that of the analyst's. Local extrapolation is significantly more prevalent when languages differ. This suggests that language barriers increase the costs of processing information, such as official disclosures, press releases, or local media, that would otherwise be more accessible to native speakers. To validate the intuition behind this result, Figure B.8 illustrates variations in language similarity across both proximate and distant countries using four benchmark cases: South Africa, Argentina, Australia, and the United Arab Emirates. For these countries, we observe both proximate and distant regions sharing the same language, ensuring variation in the data.

Economic Ties. Columns 3 and 4 of Panel B split the sample at the median level of bilateral trade intensity between the firm's and the analyst's countries. Trade intensity is measured as the bilateral imports between two countries from 2000 to 2021 (OECD data), normalized by the total trade volume of the focal country, and averaged over the sample period. Analysts are more prone to local extrapolation when economic ties are weak, consistent with the notion that stronger trade relations correlate with greater level of familiarity with the economic reality and products produced in foreign country. Figure B.8 further corroborates this by demonstrating variation in trade intensity across the same four benchmark countries, where proximate and distant regions exhibit both strong and weak trade ties.

Cultural Affinity. Columns 5 and 6 of Panel B examine the impact of cultural exposure, measured using music consumption overlap between countries. We construct an index based on Shazam's Top 200 charts for the week of December 17, 2024, normalized to range from 0 (no overlap) to 100 (perfect overlap). Higher overlap indicates greater familiarity with the cultural content of the firm's country. Local extrapolation is more frequent when cultural overlap is low, suggesting that familiarity with cultural preferences likely correlate with the greater integration and proximity with the situation in the foreign country. Figure B.8 confirms sufficient variation in cultural exposure across proximate and distant regions for the same benchmark countries.

Overall, the evidence in Table 8 extends the existing evidence on local extrapolation (e.g., Kuchler and Zafar, 2019), by documenting strong extrapolative effects across countries and economic environments, present in professional forecasters, and for long-run economic variables. These patterns appear to be largely driven by information acquisition and processing costs.

5.6 Opening the Black Box Behind Quantitative Forecasts: Qualitative Text Descriptions in Analysts' Reports

In this section, we explore the mechanisms underlying analysts' quantitative long-run expectations by examining how their expectation formation processes are shaped by differential attention to topics

relevant to firm growth prospects. Specifically, we analyze how analysts allocate attention in the qualitative sections of their reports across various topics, focusing on both the extensive margin—the decision to include a given topic—and the intensive margin—the depth of discussion devoted to covered topics. Finally, we explore how variation in topic focus contributes to heterogeneity in analysts' TGR beliefs.

5.6.1 Topical Attention Allocation

In this last part of the analysts, we explore and relate with the TGR numerical values the textual discussions taken directly from equity reports.

Topic Focus. We start with a general exploration of the topics discussed in the analyst reports to evaluate the basic properties of topic attention allocation, leveraging the LDA topic model discussed in Section 3.3. The average report focuses on few topics. Specifically, a report's top 7 topics account, on average, for 84% of the entire text content. Additional topics receive virtually no weight, with each receiving 0.4% pp of the total attention "budget" on average.

Extensive Margin Differences. Two non-mutually exclusive mechanisms could explain these patterns: (i) analysts *subjectively* focus on a limited set of topics when forming expectations, or (ii) a small number of topics *objectively* capture a firm's growth potential at a given point in time.

To distinguish between these explanations, we compare the topics discussed by pairs of analysts evaluating the same firm at the same time. If objective factors primarily drive topic selection, we would expect a high degree of overlap in the topics discussed by different analysts. Conversely, if analysts' subjective weighting of topic importance plays a larger role, we would observe greater divergence in the sets of topics covered across analysts.

To implement this exercise, we estimate the Jaccard similarity between the topics included in reports for each pair of analysts:

$$\text{Jaccard similarity} = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

where A and B are the sets of topics discussed in the reports of analysts A and B, $|A \cap B|$ represents the number of topics that overlap in both reports, and $|A \cup B|$ denotes the total number of distinct topics discussed in at least one of the reports. Simply put, Jaccard similarity quantifies the overlap between the sets of topics included in both reports, ranging from 0 (no overlap) to 1 (full overlap). In our baseline analysis, a topic is assumed to be *covered* when it accounts for at least 1% of the qualitative discussion.

For the average analyst pair, 43% of topics covered by at least one analyst of the pair are included

in both reports with an interquartile range of 17 pp (33% to 50%).[18] To make this magnitude more concrete, a Jaccard score of 40% could represent a case where analyst A focuses on topics 1, 2, 3, 4, and analyst B covers topics 1, 2, and 5. The Jaccard similarity scores indicate that analysts tend to agree on some key topics that must be included in reports, but there remain substantial differences across analysts regarding——or, put differently, person-specific factors determining——the full set of topics that are covered in a report.

Intensive Margin Differences. We next study the intensive margin of topic focus, examining the weight given to each topic that is covered in a given report. On average, analysts allocate 15% of a report's attention to each covered topic, with an interquartile range of 5 pp (10 to 20%). Thus, there exist material differences with respect to weight allocation across topics within a given report.

To further explore forecaster-specific topic focus choices at the intensive margin, we evaluate the degree of similarity in the weight of topics that overlap across reports of different forecasters. Specifically, we compare the vectors of overlapping topic weights for each pair of analysts that study the same firm at the same time, using the Cosine similarity:

$$\text{Cosine similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{4}$$

where **A** and **B** are two non-zero vectors measuring the weights allocated to overlapping topics by analysts A and B, and i denotes topic i of the overlapping vectors. Cosine similarity scores range between -1 (opposite weight) and 1 (same weight), with a value of 0 indicating no relation in topic weights.

For the average analyst pair, we obtain a Cosine similarity of 0.79 (with an interquartile range between 0.72 and 0.91), i.e., an average dissimilarity in topic weights of approximately 0.2. As a reference benchmark, Appendix Table B.1 provides the Cosine similarity distributions for overlapping topics between firms depending on whether these operate in the same or different industries. The observed average Cosine similarity of 0.79 across forecasters, while high at first glance, is only slightly higher than the average Cosine similarity of firms operating in different industries (0.74). This implies only a small incremental "convergence" of analysts forecasting the same firm at the same time in terms of chosen topic weights, relative to a cross-industry benchmark.

Overall, the results in this section indicate that the LDA model effectively identifies meaningful differences in how forecasters allocate attention to topics, both on the extensive margin (topic

[18] Importantly, these results are robust to alternative threshold values defining topic coverage. For instance, a threshold of 5%, instead of 1%, produces a Jaccard similarity score of 28%.

inclusion) and the intensive margin (topic emphasis). [19]

5.6.2 Topic Attention Differences and Variation in TGR Expectations

Having established the differences in analysts' topic focus, we now investigate the extent to which these differences predict the heterogeneity in analysts' long-term TGR expectations. This analysis bridges the quantitative and qualitative dimensions of the belief formation process.

We examine both the extensive margin of topic focus, measured by the Jaccard similarity, and the intensive margin of topic weights, measured by the Cosine similarity for overlapping topics. In essence, we ask whether differences in TGR expectations between pairs of analysts are primarily driven by differences in the topics they choose to cover or by variations in how they discuss the same topics.

Our pairwise approach leverages the unique feature of our setting: the frequent occurrence of multiple professionals evaluating the same firm at the same time. This framework allows us to explore new dimensions of attention—extensive and intensive margins—beyond the topics themselves. Additionally, it provides a novel strategy for addressing the dimensionality challenges inherent in LDA output by focusing on topics that are not systematically covered across all analysts or firms in a subgroup. This contrasts with alternative approaches in the literature, such as the use of Lasso regressions (Bybee et al., 2023a).

Panel A of Table 9 presents the results of our analysis. The dependent variable is the pairwise difference in TGR beliefs across analysts, defined as:

$$TGR_{A,B,i,t} = \frac{(TGR_{A,i,t} - TGR_{B,i,t})^2}{\left|\frac{TGR_{A,i,t}+TGR_{B,i,t}}{2}\right|}, \tag{5}$$

where A and B refer to the analysts in the pair, i indicates the firm covered in the report, and t denotes the year of the reports. Intuitively, larger values of $TGR_{A,B,i,t}$ reflect greater relative differences in analysts' TGR beliefs.

Columns (1) to (3) present results for the extensive margin, using Jaccard similarity as the variable of interest, while Columns (4) to (6) focus on the intensive margin, using Cosine similarity. Across all specifications, the coefficients on both the Jaccard and Cosine similarity measures are negative, indicating that greater similarity in the topics discussed within equity reports is associated with lower levels of disagreement and heterogeneity in analysts' TGR expectations.

Additionally, we estimate robust differences in the strength of the association between quantitative

[19] Importantly, these patterns are not driven by differences in report length across analysts. Both the Jaccard and Cosine similarity metrics normalize topic coverage as a proportion of the total report, providing unit-less measures rather than relying on word counts. Furthermore, when we restrict the sample to reports with less than a 1% difference in length, the results remain consistent, with an average Jaccard score of 43% and a Cosine similarity of 80%.

beliefs and qualitative topic discussions, depending on the extensive versus intensive margin of topic overlap. The extensive-margin Jaccard regression coefficients are negative but not statistically significant (p-values ranging from 0.21 to 0.34), indicating that we cannot reject a null relationship. By contrast, the intensive-margin Cosine point estimates are large (-0.60) and highly significant at the 5% level across all specifications. Specifically, Column (4) includes no fixed effects, Column (5) incorporates year and industry (2-digit NAICS code) fixed effects, and Column (6) uses year interacted with industry fixed effects.[20] We stress that since both the dependent variable and the variables of interest are measured as differences at the firm-year level, the analysis is naturally conducted at the within-firm-year level. Lastly, Column (7) includes both the Jaccard and Cosine similarity measures, which corroborates the conclusions from the univariate regressions in Columns (1) to (6).

Overall, these results suggest that differences in long-term expectations among professionals primarily arise from variations in how overlapping topics are prioritized, rather than from differences in the topics discussed.

5.6.3 Understanding Topic Weight Heterogeneity

In a final set of analyses, we aim to better understand the determinants and implications of differences in forecasters' intensive-margin topic focus. Additionally, we seek to validate that the topic focus results reported in Table 9 are driven by *economically intuitive* variations in forecasters' topic selection and attention.

To achieve this, we decompose the Cosine similarity measure into five broad categories: (i) macroeconomics, (ii) industry, (iii) geography, (iv) operations, and (v) valuation based on the topics labels.[21] This categorization allows us to pinpoint which types of topics are most closely associated with differences in long-term expectations.

We then reestimate three variants of the Cosine similarity score, each constructed exclusively using topics from one of three categories: macroeconomic, geographic, or industry-related topics. In Panel B of Table 9, we document a negative and statistically significant relationship between each of these category-specific Cosine similarity measures and variation in long-term beliefs. Consistent with our broader evidence on local extrapolation, the results suggest that differences in expectations are most strongly influenced by variation in attention intensity allocated to geographic and macroeconomic topics, as evidenced by their larger (absolute) regression coefficients.

For instance, the findings in Panel B of Table 9 indicate that an Argentinian analyst is more likely to provide detailed discussions on a firm's operations in South America and elaborate further on the potential effects of inflation, relative to an American analyst.

[20] The fixed effects account for broad trends occurring at the industry level over time.

[21] Appendix Table B.2 provides the complete list of LDA-derived topics along with their corresponding categories.

Finally, to further examine the role of geographic location, we analyze whether the physical distance between forecasters influences how they emphasize topics in their reports at the intensive margin. Columns (1) to (3) of Table 9, Panel C, reveal that analysts located farther apart are more likely to diverge in how they discuss the same topics.

To strengthen the link with our earlier evidence on local extrapolation presented in Table 9, Columns (4) to (6) focus specifically on analysts' geographic coverage. The results indicate that analysts situated farther away from each other allocate attention to geographic topics in systematically different ways. For example, Argentinian and Brazilian analysts are more likely to align in their discussions of a firm's geographic footprint than are American and Argentinian analysts.

Taken together, these findings suggest that local conditions and experiences influence professionals' topic focus intensity, providing a key mechanism for the observed patterns of extrapolation.

6 Conclusion

In this paper, we employ a new, large-scale dataset on terminal growth rate expectations for individual firms by professional forecasters, in which we additionally observe forecasters' qualitative discussions of their beliefs as well as their identities and individual backgrounds, to study the origins and drivers of their very long-run cash flow expectations.

We establish a series of novel findings. First, TGR expectations provide distinct economic insights relative to shorter-term forecasts, such as IBES long-term growth (LTG), as they strongly correlated with realized long-run firm growth and align with theorized drivers of long-term growth. Second, there is persistent heterogeneity in TGR expectations across forecasters, particularly for less mature firms, where country and industry factors are less influential, and forecaster demographics significantly explain this variation. Third, forecasters systematically extrapolate from the economic conditions of their country of residence when forming long-run expectations about firms in geographically distant countries—a pattern that is especially pronounced when accessing and processing local firm information is costly. Finally, differences in the thematic focus of forecasters' qualitative discussions emerge as a key mechanism underlying both long-run belief heterogeneity and extrapolation behavior.

Our findings help inform macro finance models and models of belief formation. Given the scarcity of existing evidence on truly long-term expectations, our findings are particularly informative for models that depend on distant horizons. Our findings uncover, as one key ingredient, large individual belief heterogeneity in long-run beliefs, which is potentially amplified in the long-run context, where forecasts cannot be promptly contrasted with realized outcomes, rendering them more malleable to one's experience. Another significant ingredient is extrapolation, especially extrapolation from attending to locally experienced information and signals, affecting both forecasters' qualitative belief

descriptions and their quantitative expectations.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis. 1999. High wage workers and high wage firms. *Econometrica* 67:251–333.

Andre, P., P. Schirmer, and J. Wohlfart. 2024. Mental Models of the Stock Market. *Working paper* .

Andrei, D., J. Cujean, and M. Wilson. 2023. The Lost Capital Asset Pricing Model. *The Review of Economic Studies* 90.

Asquith, P., M. B. Mikhail, and A. S. Au. 2005. Information Content of Equity Analyst Reports. *Journal of Financial Economics* 75:245–282.

Ayyagari, M., A. Demirgüç-Kunt, and V. Maksimovic. 2008. How well do institutional theories explain firms' perceptions of property rights? *The review of financial studies* 21:1833–1871.

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.

Barber, B. M., X. Huang, and T. Odean. 2016. Which Factors Matter to Investors? Evidence from Mutual Fund Flows. *The Review of Financial Studies* 29:2600–2642. URL https://doi.org/10.1093/rfs/hhw054.

Bastianello, F., P. Decaire, and M. Guenzel. 2025. Mental Models and Financial Forecasts. Tech. rep.

Ben-David, I., E. Fermand, C. M. Kuhnen, and G. Li. 2018. Expectations uncertainty and household economic behavior. Tech. rep., National Bureau of Economic Research.

Berk, J., and P. DeMarzo. 2019. *Corporate Finance*. New York.

Berk, J. B., and J. H. van Binsbergen. 2016. Assessing asset pricing models using revealed preference. *Journal of Financial Economics* 119:1–23. URL https://www.sciencedirect.com/science/article/pii/S0304405X1500149X.

Bharath, S., and T. Shumway. 2008. Forecasting Default with the Merton Distance-to-Default Model. *Review of Financial Studies* 21:1339–1369.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer. 2024a. Finance without exotic risk. Tech. rep., National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer. 2024b. Belief overreaction and stock market puzzles. *Journal of Political Economy* 132:1450–1484.

Brav, A., R. Lehavy, and R. Michaely. 2005. Using Expectations to Test Asset Pricing Models. *Financial Management* 34.

Brealey, R., S. Myers, and A. Marcus. 2022. *Principles of Corporate Finance*. New York.

Bybee, L., B. Kelly, A. Manela, and D. Xiu. 2023a. Business News and Business Cycles. *Forthcoming at the Journal of Finance* .

Bybee, L., B. Kelly, and Y. Su. 2023b. Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text. *The Review of Financial Studies* 36:4759–4787. URL https://doi.org/10.1093/rfs/hhad042.

Charles, C., and C. Kendall. 2024. Causal Narratives. *Working paper* .

Charles, C., and P. Sui. 2024. Marketwide Memory. *Available at SSRN* .

Chaudhry, A. 2025. The Impact of Prices on Analyst Cash Flow Expectations: Reconciling Subjective Beliefs Data with Rational Discount Rate Variation. *Journal of Financial Economics, Conditionally Accepted* .

Chen, L., Z. Da, and X. Zhao. 2013. What Drives Stock Price Movements? *The Review of Financial Studies* 26:841–876.

Coibion, O., and Y. Gorodnichenko. 2012. What can survey forecasts tell us about information rigidities? *Journal of Political Economy* 120:116–159.

D'Acunto, F., P. Ghosh, and A. G. Rossi. 2022a. How costly are cultural biases? Evidence from FinTech .

D'Acunto, F., D. Hoang, M. Paloviita, and M. Weber. 2023. IQ, expectations, and choice. *The Review of Economic Studies* 90:2292–2325.

D'Acunto, F., J. Xie, and J. Yao. 2022b. Trust and contracts: Empirical evidence. Tech. rep., LawFin Working Paper.

Das, S., C. M. Kuhnen, and S. Nagel. 2020. Socioeconomic status and macroeconomic expectations. *The Review of Financial Studies* 33:395–432.

De La O, R., X. Han, and S. Myers. 2024. The Cross-section of Subjective Expectations: Understanding Prices and Anomalies. *Working Paper* .

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

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

Decaire, P. H., D. Sosyura, and M. Wittry. 2023. Model Input Choice and Valuation Disagreement. *Working paper* .

Demgensky, L., and U. Fritsche. 2023. Narratives on the causes of inflation in Germany: First results of a pilot study. Tech. rep.

D'Acunto, F., U. Malmendier, J. Ospina, and M. Weber. 2021. Exposure to grocery prices and inflation expectations. *Journal of Political Economy* 129:1615–1639.

D'Acunto, F., and M. Weber. 2024. Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics* 16.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus. 2021. Five Facts about Beliefs and Portfolios. *American Economic Review* 111:1481–1522.

Gilardi, F., M. Alizadeh, and M. Kubli. 2023. ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks. *Working paper: arXiv:2303.15056* .

Golez, B. 2014. Expected returns and dividend growth rates implied by derivative markets. *The Review of Financial Studies* 27:790–822.

Graham, J. R., and C. Harvey. 2001. The Theory and Practice of Corporate Finance: Evidence from the Field. *Journal of Financial Economics* 60:187–243.

Guenzel, M., and U. Malmendier. 2020. Behavioral Corporate Finance: The Life Cycle of a CEO Career. *Oxford Research Encyclopedia of Economics and Finance, September 2020* .

Guercio, D. D., and P. A. Tkac. 2002. The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds. *The Journal of Financial and Quantitative Analysis* 37:523–557. URL http://www.jstor.org/stable/3595011.

Hartzmark, S. M., S. D. Hirshman, and A. Imas. 2021. Ownership, Learning, and Beliefs*. *The Quarterly Journal of Economics* 136:1665–1717. URL https://doi.org/10.1093/qje/qjab010.

Hashim, N. A., and N. C. Strong. 2018. Do Analysts' Cash Flow Forecasts Improve Their Target Price Accuracy? *Contemporary Accounting Research* 35:1816–1842.

Hillenbrand, S., and O. McCarthy. 2024. Valuing Stocks With Earnings. *Working Paper* .

Jones, K. 1996. Trust as an Affective Attitude. *Ethics* 107:4–25.

Kendall, C., and R. Oprea. 2023. On the Complexity of Forming Mental Models. *Working paper* .

Klepper, S. 1996. Entry, exit, growth, and innovation over the product life cycle. *The American Economic Review* pp. 562–583.

Kuchler, T., and B. Zafar. 2019. Personal experiences and expectations about aggregate outcomes. *The Journal of Finance* 74:2491–2542.

Malmendier, U. 2018. Behavioral Corporate Finance. In *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 1, pp. 277–379. Elsevier.

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

Malmendier, U., and S. Nagel. 2016. Learning from inflation experiences. *The Quarterly Journal of Economics* 131:53–87.

Mukhlynina, L., and K. G. Nyborg. 2018. The Choice of Valuation Techniques in Practice: Education versus Profession. *Working Paper* .

Nagel, S. 2021. Empirically Grounded Models of Subjective Beliefs, in: Perspectives on the future of asset pricing. *The Review of Financial Studies* 34:2126–2160.

Nagel, S. 2024. Stock Market Valuation: Explanations, Non-Explanations, and Some Open Questions. Comments Given at the NBER Summer Institute Asset Pricing Group Meeting (Expanded Version) meeting.

Nagel, S., and Z. Xu. 2022. Asset Pricing with Fading Memory. *Review of Financial Studies* 35:2190–2245.

Porta, R. L. 1996. Expectations and the Cross-Section of Stock Returns. *Journal of Finance* 51:1715–1742.

Pruitt, S. 2024. Dogs and Cats Living Together: A Defense of Cash-Flow Predictability. *Working Paper* .

Sabbatucci, R. 2022. Are dividends and stock returns predictable? New evidence using M&A cash flows. *Working Paper* .

Weber, M., F. D'Acunto, Y. Gorodnichenko, and O. Coibion. 2022. The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *Journal of Economic Perspectives* 36:157–184.

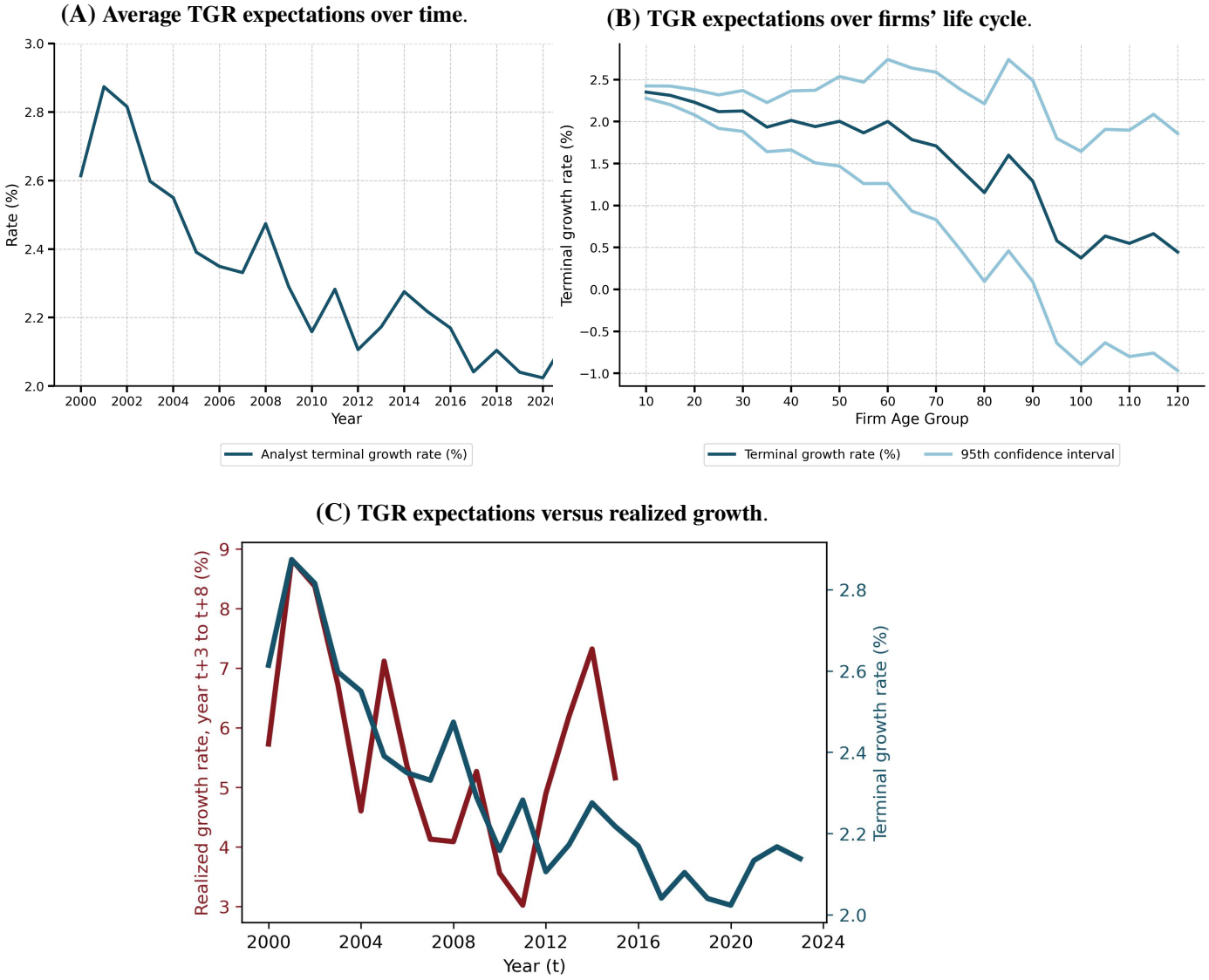


Figure 1: TGR expectations Panel A figure plots the time trend of average terminal growth rates across analysts and firms over the sample period from 2000 to 2023. Panel B figure plots analyst terminal growth rates over firms' life cycle. The solid blue line corresponds to Age group fixed effects obtained from the regression: $TGR_{i,j,t} = k + I_{\text{Age group}} + I_{\text{Firm}} + I_{\text{Year}} + \epsilon_{i,j,t}$, for which we add back the regression constant (k) to better illustrate the terminal growth rate level. The pale blue lines correspond to the $I_{\text{Age group}}$ regression estimate 95^{th} confidence interval, for which we add back the regression constant (k) for consistency with the other lines of the figure, and illustration purposes. The variable Firm Age Group captures firm age groups measured in 5-year increments. For example, Firm Age Group = 10 corresponds to firms that are 5 to 10 years old. Panel C Figure compares the time trends of ex-post realized earnings growth at the $t+3$ to $t+8$ horizon with analyst terminal growth rates. Realized earnings growth corresponds to the geometric average realized from $t+3$ to $t+8$, averaged by year (t) in the sample. Terminal growth rate is the average terminal growth rate by year. The x-axis denotes years. The left-hand-side y-axis is associated with the ex-post realized returns, and the right-hand-side y-axis relates to the terminal growth rate.

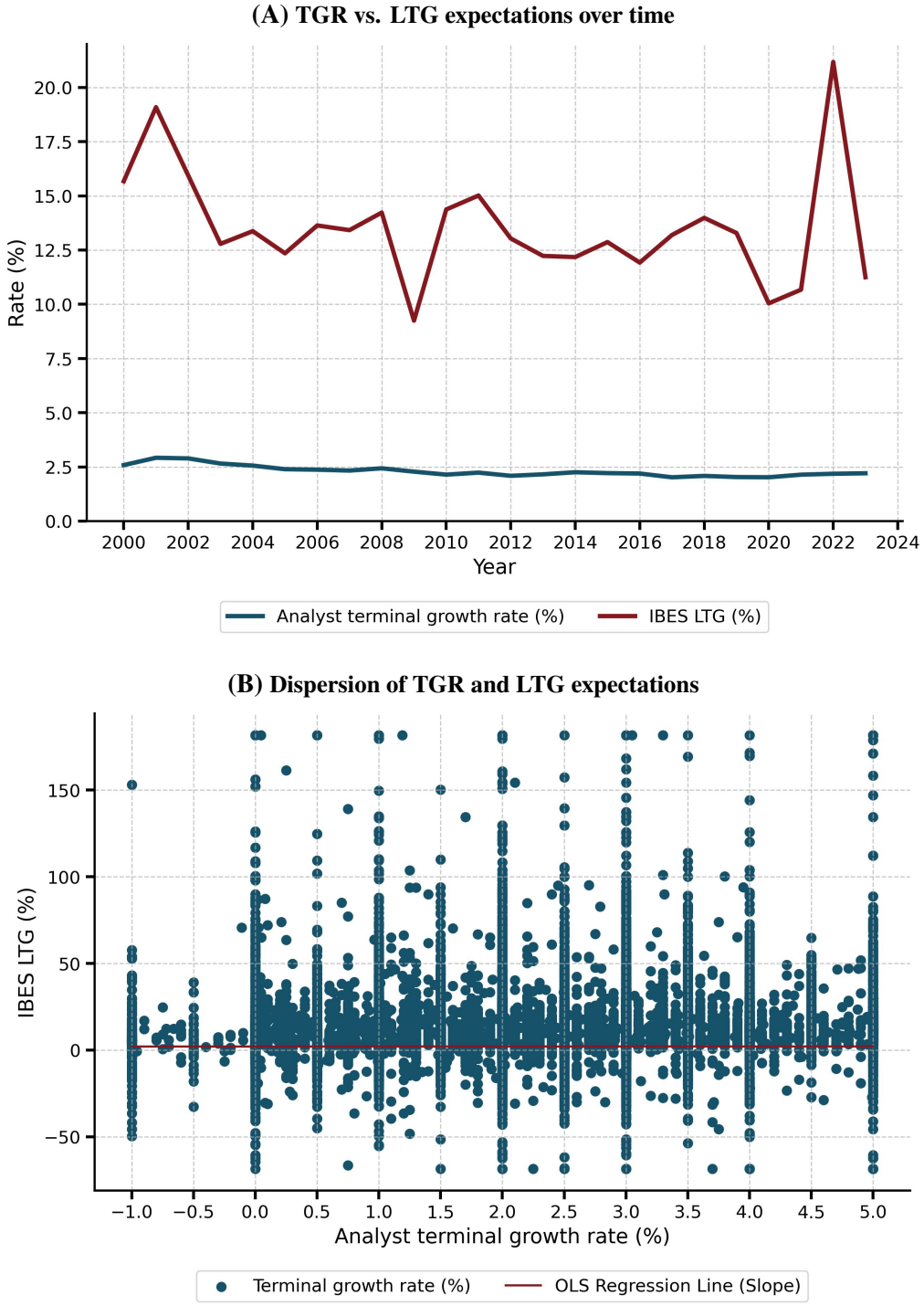


Figure 2: TGR versus IBES LTG expectations For Panel A, the x-axis is expressed in years, and the y-axis in percentage points. Panel A compares the time trend of IBES LTG expectations with analysts' terminal growth rates over the sample period. Panel B's x-axis represents analysts' terminal growth rates, and the y-axis corresponds to the firm-year LTG consensus (median) provided by IBES. Each blue dot denotes an observation. The red line is the slope from an OLS regression between IBES LTG and terminal growth rates. For illustration purposes and readability, we winsorize the IBES LTG variable at the 0.1 and 99.9^{th} percentiles.

Table 1: Summary statistics This table reports summary statistics. The sample consists of 8,010 firms with observations from 41,429 equity reports in 2000–2023. Panel A describes the sample coverage. Panel B focuses on DCF variables associated with equity reports, Panel C provides information about the analysts included in the analysis, and Panels D and E examine economic data series. Variables definitions appear in Appendix A.

Panel A: Firms and Coverage						
		No. Firm			No. Obs	
Terminal growth rate sample		8,010			41,429	
Equity Analyst					5,246	
Equity Analyst (LinkedIn Matched)					2,565	
Brokerage house					42	
Panel B: Equity Reports	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Foreign reports	0.31	0.00	0.00	1.00	0.46	41,129
No. reports by country	845.49	11.00	240.00	581.00	2,140.80	49
Analyst terminal growth rate $(TGR)_{i,j,t}$ (%)	2.22	1.50	2.00	3.00	1.31	41,429
$TGR_{i,j,t}$ (%) (IBES Overlap)	2.21	1.50	2.00	3.00	1.29	25,615
IBES $LTG_{i,j,t}$ (%)	13.03	5.50	10.30	17.00	22.48	25,615
Correlation(IBES LTG, TGR)	0.07^{***}					
Panel C: Analyst Information	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Female	0.16	0.00	0.00	0.00	0.36	2,565
Asian	0.20	0.00	0.00	0.00	0.40	2,565
Black	0.03	0.00	0.00	0.00	0.16	2,565
Caucasian	0.63	0.00	1.00	1.00	0.48	2,565
Hispanic	0.04	0.00	0.00	0.00	0.18	2,565
Master degree	0.27	0.00	0.00	1.00	0.44	2,565
MBA	0.21	0.00	0.00	0.00	0.41	2,565
Birth year	1977.56	1971.00	1978.00	1985.00	9.41	1,544
No. Reports by Analyst	7.90	1.00	3.00	7.00	16.29	5,246
No. analyst by country	114.04	7.00	39.00	91.00	252.61	46
No. analyst by brokerage house	138.76	22.00	42.50	180.00	184.69	42
Panel D: Economic Data Series—All sample	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Firm country						
Real GDP $growth_{i,j,t}$ (%)	2.55	1.55	2.29	3.74	3.36	38,370
$Inflation_{i,j,t}$ (%)	2.57	1.26	2.01	3.16	2.50	38,247
10-year Treasury $yield_{i,j,t}$ (%)	3.15	1.68	2.91	4.10	2.30	40,000
Panel E: Economic Data Series—Foreign reports	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Analyst country						
Real GDP $growth_{i,j,t}$ (%)	2.01	1.45	2.29	3.12	3.70	12,296
$Inflation_{i,j,t}$ (%)	2.49	1.45	2.27	2.85	1.93	12,285
10-year Treasury $yield_{i,j,t}$ (%)	3.02	1.51	2.92	4.14	2.10	12,828
Firm country						
Real GDP $growth_{i,j,t}$ (%)	2.11	0.96	2.21	3.82	3.58	11,802
$Inflation_{i,j,t}$ (%)	2.43	0.90	1.83	3.06	2.75	11,686
10-year Treasury $yield_{i,j,t}$ (%)	3.02	0.97	2.75	4.11	2.76	11,830

Table 2: TGR, cost of equity, and realized returns Panel A studies the relation between the Terminal growth rate (TGR) and realized 1-year total equity returns. Realized total Equity Returns is taken from Refinitiv Eikon. TGR is taken from equity analysts report. Subjective Cost of Equity is obtain by computing rf+$\beta * ERP$, all three subjective inputs taken directly from the equity reports. Panel B investigate the relation between the subjective cost of equity, and TGR. Variable definitions appear Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Panel A: Realized Returns	**Realized Total Equity Returns (%)$_{j,t+1}$**		
	(1)	(2)	(3)
Subjective Cost of Equity (%)$_{i,j,t}$	1.70***	0.95**	0.92*
	(0.36)	(0.46)	(0.50)
TGR (%)$_{i,j,t}$	-0.88*	-2.14***	-1.71**
	(0.45)	(0.71)	(0.73)
Year FE	No	Yes	No
Firm FE	No	Yes	Yes
Firm Country * Year FE	No	No	Yes
Observations	5,579	4,476	4,273
F Statistics	11.69	6.86	4.45
R^2	0.01	0.48	0.59
Panel B: TGR and the Cost of Equity	**TGR$_{i,j,t}$**		
	(1)	(2)	(3)
Subjective Cost of Equity (%)$_{i,j,t}$	0.08***	-0.04	-0.14
	(0.01)	(0.04)	(0.10)
Year FE	No	No	No
Firm FE	No	Yes	No
Firm Country * Year FE	No	Yes	Yes
Firm * Analyst FE	No	No	Yes
Observations	8,574	6,657	4,168
F Statistics	51.33	0.83	1.94
R^2	0.01	0.59	0.64

Table 3: Economic content of TGR estimates In Panel A, this table decomposes ex-post realized earnings growth at different horizons based on analysts' terminal growth rates, and IBES LTG. The unit of observation is at the firm *i*, analyst *j*, and forecast year *t* levels. The sample period is 2000–2023. The dependent variables in Columns 1 to 6 are the geometric average of firms' earnings annual growth rates experienced by firm *i* for the 5-year periods *t+n* to *t+n+4*, for different values of *n*. *Terminal growth rate*$_{i,j,t}$ is the terminal growth rate expectation by the analyst, measured in percentage points. $LTG_{j,t}$ is IBES median concensus LTG in percentage points measured at time *t* for firm *i*. In Panel B, we explore the relation between default risk, measured using the Merton's Model adjustment from Bharath and Shumway (2008), firm age, and the TGR. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Panel A: Realized Earnings Growth	$\Pi_{t+1}^{t+5}(1+g_{j,t})^{1/5}-1$		$\Pi_{t+3}^{t+8}(1+g_{j,t})^{1/5}-1$		$\Pi_{t+3}^{t+13}(1+g_{j,t})^{1/10}-1$	
	(1)	(2)	(3)	(4)	(5)	(6)
Terminal growth rate (%)$_{i,j,t}$	0.69***	0.55***	0.66***	0.46***	0.53***	0.40***
	(0.13)	(0.15)	(0.14)	(0.16)	(0.11)	(0.11)
IBES Median Concensus LTG (%)$_{j,t}$	0.30***	0.32***	0.08***	0.08***	0.05***	0.04**
	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Firm Country * Industry * Year FE	No	Yes	No	Yes	No	Yes
Observations	15,441	14,017	12,735	11,506	7,225	6,481
F-statistics	75.83	45.12	21.37	10.31	16.85	8.48
R^2	0.10	0.41	0.01	0.35	0.02	0.42

Panel B: TGR, Default, and Firm Age	Terminal growth rate (%)$_{i,j,t}$				
	(1)	(2)	(3)	(4)	(5)
Default risk (10-year horizon)$_{i,j,t}$	-0.19***			-0.25***	
	(0.04)			(0.07)	
Default risk (20-year horizon)$_{i,j,t}$		-0.20***			-0.26***
		(0.03)			(0.06)
ln(Firm post-IPO age)$_{i,j,t}$			-0.16***	-0.25***	-0.25***
			(0.04)	(0.05)	(0.05)
Firm Country * Industry * Year FE	Yes	Yes	Yes	Yes	Yes
Observations	26,172	26,172	22,124	14,745	14,745
F-statistics	29.10	33.70	18.16	10.64	12.16
R^2	0.19	0.19	0.17	0.19	0.19

Table 4: ANOVA variance decomposition This table presents a variance decomposition of analyst terminal growth rates using ANOVA. For this analysis, we exclude from the sample firms, analysts, and brokerage houses with fewer than 2 observations in total, as for these observations, their respective fixed effects fully explain the associated variation in terminal growth rates. We also exclude brokerage houses that are not associated with equity analysts that switched between employers during the sample period (2000–2023), since it is not possible to distinguish between the effect of those brokerage houses and their analysts (Abowd et al., 1999). Panel A includes this entire subsample. Panel B excludes firms with fewer than 10 observations. To account for the fact that the sample used to conduct the analysis is unbalanced, as we do not observe a terminal growth rate estimate for every analyst-firm-year pair during the entire sample period, we estimate the sequential partial sum of squares. The adjusted partial R^2 is defined as: $1-(1-(\frac{Pss_i}{TSS})*\frac{N-1}{N-df_i-1})$, where i denotes a specific set of fixed effects (brokerage house, analyst, or firm). In both Panels, Columns 1-4 perform the variance decomposition for the TGR variable, and Column 5-8 apply it to the IBES 1-Year EPS growth forecast. The IBES sample used for the decomposition covers the period 2000-2023, and includes the firms that are also covered in the TGR to allow for a direct comparison between both set of results. The IBES 1-Year EPS growth forecast, is measured using the earliest forecast made by an analysts for a given fiscal year, and must be at least produced by the analyst 9-month before the target date, and at most 1 year. To measure the analyst expected growth, we divide the EPS forecast by the lastest realized EPS of the firm observed by the analyst at the time of the forecast, using the data provided in IBES (ACTUAL). Ultimately, *1-Year Expected EPS Growth Rate* is equal to VALUE/ACTUAL-1, whenever a growth rate can be measured using that data. Variable definitions appear in Appendix A.

Panel A: Anova

	TGR				1-Year Expected EPS Growth Rate			
	All sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$	All sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brokerage house FE								
Share of total R^2 (%)	3	3	4	6	1	2	2	3
Degrees of Freedom	39	38	35	27	421	221	102	44
Adjusted Partial R^2	0.02	0.2	0.02	0.02	-0.01	-0.01	-0.01	-0.00
Analyst FE								
Share of total R^2 (%)	55	48	37	23	33	24	7	1
Degrees of Freedom	3,522	1,038	397	108	3,569	736	149	30
Adjusted Partial R^2	0.22	0.21	0.17	0.09	-0.02	0.01	-0.00	-0.01
Firm FE								
Share of total R^2 (%)	41	49	59	72	66	74	91	96
Degrees of Freedom	5,091	4,570	3,593	2,220	2,608	2,207	1,446	740
Adjusted Partial R^2	0.09	0.10	0.13	0.11	0.15	-0.00	0.11	0.13
Total R^2	0.53	0.50	0.50	0.45	0.36	0.20	0.34	0.39
Observations	36,889	27,737	18,808	9,261	26,152	15,113	6,302	2,609

Panel B: Anova—Obs per firm ≥ 10

	TGR				1-Year Expected EPS Growth Rate			
	All sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$	All sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brokerage house FE								
Share of total R^2 (%)	4	6	8	14	5	2	3	15
Degrees of Freedom	38	37	34	27	356	212	100	44
Adjusted Partial R^2	0.02	0.03	0.05	0.08	-0.01	-0.01	-0.01	0.01
Analyst FE								
Share of total R^2 (%)	85	74	67	55	56	38	12	1
Degrees of Freedom	2,709	963	386	107	3,087	728	148	30
Adjusted Partial R^2	0.51	0.40	0.39	0.33	-0.04	0.00	-0.00	-0.01
Firm FE								
Share of total R^2 (%)	12	21	25	32	39	60	85	84
Degrees of Freedom	1,187	1,177	1,141	866	1,179	1,158	904	488
Adjusted Partial R^2	0.03	0.06	0.06	0.06	0.03	0.00	0.04	-0.09
Total R^2	0.68	0.58	0.61	0.63	0.22	0.17	0.25	0.20
Observations	22,122	17,278	12,030	5,814	19,470	11,933	5,047	2,093

Table 5: TGR expectations over firms' life cycle This table evaluates the share of the second-stage dependent variable explained by the right-hand-side explanatory variables over firm life cycle. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023.Column 1 reports the R^2 obtained from estimating the Case 1 Stage 2 regression for the subset of firms whose age is within the range shown in the first column. Column 2 reports the share of the Case 1 Stage 2 regression R^2 that can be attributed to the *Firm country* and *Industry* factors. We obtain this estimate by performing a sequential ANOVA decomposition for the specification shown in Case 1 Stage 2. Column 3 reports the R^2 obtained from estimating the Case 2 Stage 2 regression for the subset of firms whose age is within the range shown in the first column. Column 4 performs a conservative back-of-the-envelope calculation to estimate the share of TGR in Case 1 Stage 1 that can be explained by the firm factors, after controlling for two firm characteristics: firm country and industry. Specifically, we obtain that magnitude by computing $(1 - R^2_{Column\ 1}) \times \phi$. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares.

	Case 1		Case 2	
Stage 1:	None		$TGR_{i,j,t} = I_{\text{Firm country}} + I_{\text{Industry}} + I_{\text{Analyst}} + I_{\text{Brokerage}} + \epsilon_{i,j,t}$	
Stage 2:	$TGR_{i,j,t} = I_{\text{Firm country}} + I_{\text{Industry}} + I_{\text{Analyst}} + I_{\text{Brokerage}} + \epsilon_{i,j,t}$		$\hat{\epsilon}_{i,j,t} = I_{\text{Firm}} + \upsilon_{i,j,t}$	
	Stage 2 R^2	**Share of R^2 Driven by $I_{\text{Firm country}} + I_{\text{Industry}}$**	**Stage 2 R^2 (ϕ)**	**Share of $TGR_{i,j,t}$ Driven by Firm Residual Factors**
	(1)	(2)	(3)	(4)
0 <= Firm age < 5	0.38	7%	0.12	7%
5 <= Firm age < 10	0.69	13%	0.37	12%
10 <= Firm age < 15	0.75	12%	0.31	8%
15 <= Firm age < 20	0.69	22%	0.28	9%
20 <= Firm age < 25	0.68	23%	0.27	9%
25 <= Firm age < 30	0.76	30%	0.26	6%
30 <= Firm age < 35	0.68	24%	0.13	4%
35 <= Firm age < 40	0.80	26%	0.22	4%
40 <= Firm age < 45	0.78	31%	0.10	2%
45 <= Firm age < 50	0.79	39%	0.00	0%
50 <= Firm age < 55	0.89	49%	0.13	1%
55 <= Firm age < 60	0.86	62%	0.05	1%
60 <= Firm age < 65	0.93	57%	0.00	0%
65 <= Firm age < 70	0.93	59%	0.01	0%
70 <= Firm age < 75	0.87	71%	0.11	1%
75 <= Firm age < 80	0.61	37%	0.13	5%
80 <= Firm age < 85	0.91	64%	0.01	0%
85 <= Firm age < 90	0.90	31%	0.01	0%
90 <= Firm age < 95	0.96	77%	0.00	0%

Table 6: TGR expectation dispersion This table presents the results of the dispersion of terminal growth rate expectations over firms' life cycle. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023. The dependent variable corresponds to the difference between the analyst's estimate of the terminal growth rate and the industry (k) consensus for the firm age group in the forecast year t squared and scaled by the industry average on that year. The dependent variables are indicator variables equal to one if the firm age lies within the range associated with the indicator variable and 0 otherwise. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Terminal Growth Rate dispersion	$(TGR_{i,j,t} - \overline{TGR}_{k,t}^{\text{Age Group}})^2/\overline{TGR}_{k,t}$		
	(1)	(2)	(3)
5 <= Firm age < 10$_{i,j,t}$	-0.01	-0.01	-0.03
	(0.03)	(0.02)	(0.03)
10 <= Firm age < 15$_{i,j,t}$	-0.04	-0.03	-0.05*
	(0.03)	(0.03)	(0.03)
15 <= Firm age < 20$_{i,j,t}$	-0.08**	-0.08***	-0.10***
	(0.03)	(0.03)	(0.03)
20 <= Firm age < 25$_{i,j,t}$	-0.06*	-0.07**	-0.10***
	(0.03)	(0.03)	(0.04)
25 <= Firm age < 30$_{i,j,t}$	-0.04	-0.06	-0.09**
	(0.04)	(0.04)	(0.04)
30 <= Firm age < 35$_{i,j,t}$	-0.17***	-0.20***	-0.26***
	(0.04)	(0.04)	(0.04)
35 <= Firm age < 40$_{i,j,t}$	-0.27***	-0.30***	-0.38***
	(0.04)	(0.04)	(0.04)
40 <= Firm age < 45$_{i,j,t}$	-0.32***	-0.33***	-0.38***
	(0.05)	(0.05)	(0.06)
45 <= Firm age < 50$_{i,j,t}$	-0.42***	-0.40***	-0.46***
	(0.05)	(0.05)	(0.05)
50 <= Firm age < 55$_{i,j,t}$	-0.37***	-0.37***	-0.39***
	(0.06)	(0.07)	(0.07)
55 <= Firm age < 60$_{i,j,t}$	-0.45***	-0.47***	-0.54***
	(0.05)	(0.06)	(0.07)
60 <= Firm age < 65$_{i,j,t}$	-0.26***	-0.34***	-0.41***
	(0.09)	(0.09)	(0.10)
65 <= Firm age < 70$_{i,j,t}$	-0.54***	-0.52***	-0.58***
	(0.06)	(0.05)	(0.05)
70 <= Firm age < 75$_{i,j,t}$	-0.28***	-0.35***	-0.45***
	(0.10)	(0.09)	(0.10)
75 <= Firm age < 80$_{i,j,t}$	-0.29***	-0.36***	-0.24
	(0.09)	(0.11)	(0.15)
80 <= Firm age < 85$_{i,j,t}$	-0.46***	-0.49***	-0.54***
	(0.09)	(0.13)	(0.15)
85 <= Firm age < 90$_{i,j,t}$	-0.54***	-0.43***	-0.46***
	(0.07)	(0.09)	(0.09)
90 <= Firm age < 95$_{i,j,t}$	-0.30*	-0.36***	-0.41***
	(0.16)	(0.12)	(0.13)
95 <= Firm age < 100$_{i,j,t}$	-0.61***	-0.60***	-0.64***
	(0.03)	(0.05)	(0.06)
Firm Size$_{i,t}$			0.03***
			(0.01)
TGR$_{i,j,t}$			0.02
			(0.02)
Constant	0.70***	0.70***	0.08
	(0.02)	(0.02)	(0.12)
Year*Industry FE	No	Yes	Yes
H_0 = All Age coefficients are the same (Bonferroni p-value)	0.00	0.00	0.00
Observations	24,280	24,255	21,295
F Statistics	37.53	19.07	16.57
R^2	0.01	0.08	0.08

Table 7: Analyst fixed effects decomposition This table shows the combined explanatory power of various fixed effects for terminal growth rates and presents a variance decomposition of analyst fixed effects in terminal growth rate expectations. For this analysis, we exclude from the sample firms, analysts, and brokerage houses with fewer than 2 observations in total, as for these observations, their respective fixed effects fully explain the associated variation in terminal growth rates. We also exclude brokerage houses that are not associated with equity analysts that switched between employers during the sample period (2000–2023), since it is not possible to distinguish between the effect of those brokerage houses and their analysts (Abowd et al., 1999). In column 1, we use the entire subsample; column 2 restricts the analysis to firms that have both domestic and foreign coverage, and column 3 focuses on the foreign coverage for the firms included in column 2. Panel A shows the combined explanatory power of three primitive variables—firm, brokerage house, and analyst—for the terminal growth rate. The unit of observation is at firm i, analyst j, and forecast year t levels. Panel B performs a variance decomposition of the analyst fixed effects into 5 dependent variables using ANOVA. Gender is a binary variable equal to 1 if the analyst is female and 0 otherwise. Race is a categorical variable indicating White, Asian, Hispanic, and other races. Education is a binary variable equal to 1 if the analyst has a graduate degree and 0 otherwise. Birthyear is a series of birth year indicators. To account for the fact that the sample used to conduct the analysis is unbalanced, as we do not observe a terminal growth rate estimate for every analyst-firm-year pair during the entire sample period, we estimate the sequential partial sum of squares. The adjusted partial R^2 is defined as: $1 - (1 - (\frac{Pss_i}{TSS}) * \frac{N-1}{N-df_i-1})$, where i denotes a specific set of fixed effects (brokerage house, analyst, or firm). Variable definitions appear in Appendix A.

Panel A: Fixed effects	Terminal growth rate$_{i,j,t}$					
Sample filters	At least 2 observations per firm, and brokerage house, and analyst			At least 5 observations per firm, and brokerage house, and analyst		
	All Sample	Has foreign coverage	Exclude domestic coverage	All Sample	Has foreign coverage	Exclude domestic coverage
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.17***	2.07***	2.11***	2.14***	2.06***	2.09***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes	Yes	Yes
Analyst FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,768	5,624	3,616	10,261	4,754	3,139
R^2	0.77	0.86	0.74	0.64	0.68	0.73
Panel B: ANOVA decomposition	**Analyst FE**					
Sample filters	At least 2 observations per firm, and brokerage house, and analyst			At least 5 observations per firm, and brokerage house, and analyst		
	All Sample	Has foreign coverage	Exclude domestic coverage	All Sample	Has foreign coverage	Exclude domestic coverage
	(1)	(2)	(3)	(4)	(5)	(6)
Gender FE						
Share of total R^2 (%)	0	1	0	1	0	0
Degrees of Freedom	1	1	1	1	1	1
Adjusted Partial R^2	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
Education FE						
Share of total R^2 (%)	15	1	1	4	1	1
Degrees of Freedom	2	2	2	2	2	2
Adjusted Partial R^2	-0.00	-0.00	-0.00	0.01	-0.00	-0.00
Race FE						
Share of total R^2 (%)	9	2	2	2	1	1
Degrees of Freedom	3	3	3	3	3	3
Adjusted Partial R^2	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
Birthyear FE						
Share of total R^2 (%)	56	38	38	28	38	48
Degrees of Freedom	47	43	43	44	42	48
Adjusted Partial R^2	0.04	0.01	0.00	-0.00	0.01	0.00
Analyst country FE						
Share of total R^2 (%)	58	58	59	65	60	49
Degrees of Freedom	35	36	28	32	33	26
Adjusted Partial R^2	0.02	0.07	0.11	0.12	0.10	0.06
Total R^2	0.16	0.22	0.30	0.25	0.28	0.29
Observations	884	576	391	602	427	308

Table 8: Analyst–firm distance, analyst local economic conditions, and extrapolation This table explores the effect of analyst local economic conditions on their terminal growth rate expectations. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023. The dependent variable, *Terminal growth rate*$_{i,j,t}$, is equal to the terminal growth rate expectation by the analyst, measured in percentage points. The 10-year average real GDP growth series, for both the firm and the analyst's country, are taken from the World Bank data series. They are measured as the trailing 10-year average. The 10-year average inflation rates, for both the firm and the analyst's country, are constructed from data from the World Bank. They are measured as the trailing 10-year average. *Far* is a binary variable equal to 1 if the geographic distance between the analyst's country and the firm's country is above the sample median, and 0 otherwise. We measure the distance between countries as the shortest distance between the two countries' capitals. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Panel A: Local Extrapolation	Terminal growth rate$_{i,j,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
10-year avg. real GDP growth (Firm)$_{i,j,t}$	0.13***	0.09***	0.09***			
	(0.02)	(0.03)	(0.03)			
10-year avg. real GDP growth (Analyst)$_{i,j,t}$		0.09***	-0.04			
		(0.03)	(0.05)			
10-year avg. real GDP growth (Analyst)$_{i,j,t} * I^{far}_{i,j,t}$			0.17***			
			(0.06)			
10-year avg. inflation rate (Firm)$_{i,j,t}$				0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
10-year avg. inflation rate (Analyst)$_{i,j,t}$					0.00	-0.00
					(0.01)	(0.00)
10-year avg. inflation rate (Analyst)$_{i,j,t} * I^{far}_{i,j,t}$						0.05
						(0.04)
Analyst country*firm country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,692	11,675	11,675	11,592	11,575	11,575
F Statistics	38.74	28.66	20.74	0.87	0.52	0.81
R^2	0.05	0.05	0.05	0.05	0.05	0.05

Panel B: Drivers of Local Extrapolation	Terminal growth rate$_{i,j,t}$					
	Language		Trade relation		Cultural affinity	
	Different	Same	< median	≥ median	< median	≥ median
	(1)	(2)	(3)	(4)	(5)	(6)
10-year avg. real GDP growth (Analyst)$_{i,j,t} * I^{far}_{i,j,t}$	0.16***	0.12	0.35***	0.11	0.22***	0.02
	(0.06)	(0.21)	(0.09)	(0.07)	(0.07)	(0.09)
Analyst country*firm country FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	10-year avg. real GDP growth (Firm)$_{i,j,t}$, 10-year avg. real GDP growth (Analyst)$_{i,j,t}$					
Observations	9,738	1,937	5,794	5,840	5,585	5,579
F Statistics	18.23	5.51	13.68	11.56	14.75	11.69
R^2	0.05	0.04	0.04	0.12	0.17	0.03

Table 9: Variation in TGR Expectations and Differences in Qualitative Discussion of Topics This table explores the relation between differences in textual discussion and disagreement in terminal growth rate expectations between pairs of analysts. The unit of observation is at the analyst a, analyst b, firm i, and forecast year t levels. The sample period is 2000–2023. The dependent variable, $Disagreement_{a,b,i,t}$, is equal to the difference between analyst A and analyst B TGR estimates squared and scaled by the absolute value of their sum: $(TGR_{a,i,t} - TGR_{b,i,t})^2/|TGR_{a,i,t} + TGR_{b,i,t}|$. The variables of interest are Jaccard and Cosine scores, which respectively measure the similarity levels between the two analysts' textual valuation theses on the extensive and intensive margins, respectively. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the analyst pair level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Panel A: Differences in Qualitative Discussion	$Disagreement_{a,b,i,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Jaccard $_{a,b,i,t}$	-0.27	-0.26	-0.35				-0.40
	(0.26)	(0.27)	(0.28)				(0.38)
Cosine $_{a,b,i,t}$				-0.61**	-0.56**	-0.60**	-0.59**
				(0.24)	(0.24)	(0.26)	(0.25)
Year FE	No	Yes	No	No	Yes	No	No
Industry FE	No	Yes	No	No	Yes	No	No
Industry*Year FE	No	No	Yes	No	No	Yes	Yes
Observations	31,828	31,785	31,765	18,883	18,863	18,836	18,836
F Statistics	1.06	0.92	1.54	6.48	5.33	5.39	2.71
R^2	0.00	0.01	0.05	0.00	0.02	0.07	0.07

Panel B: Category-Specific Topic Weights	$Disagreement_{a,b,i,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Cosine $^{Macro}_{a,b,i,t}$	-1.32**	-1.29**				
	(0.61)	(0.60)				
Cosine $^{Industry}_{a,b,i,t}$			-0.97**	-0.90*		
			(0.46)	(0.48)		
Cosine $^{Geography}_{a,b,i,t}$					-2.16***	-1.65**
					(0.80)	(0.79)
Industry*Year FE	No	Yes	No	Yes	No	Yes
Observations	11,644	11,578	17,688	17,637	11,053	11,008
F Statistics	4.74	4.54	4.40	3.55	7.27	4.32
R^2	0.00	0.07	0.00	0.06	0.00	0.09

Panel C: Distance Between Analysts	$Cosine_{a,b,i,t}$			$Cosine^{Geography}_{a,b,i,t}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$I^{far}_{a,b,t}$	-0.02***	-0.02***	-0.02***	-0.01***	-0.01***	-0.01***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Year FE	No	Yes	No	No	Yes	No
Industry FE	No	Yes	No	No	Yes	No
Industry*Year FE	No	No	Yes	No	No	Yes
Observations	22,463	22,462	22,434	18,156	18,155	18,124
F Statistics	10.39	15.19	17.57	29.08	17.59	10.55
R^2	0.00	0.03	0.07	0.00	0.02	0.06

Appendix A Variable Definitions

Table A.1: Variable Definitions

Subscript t indicates the forecast year, i indicates a firm, j indicates a contributor, and c identifies a country.

Variable	Definition
10-year average inflation rate (Analyst)$_{i,j,t}$	The 10-year rolling average of the analyst's country inflation rate, obtained from the World Bank.
10-year average inflation rate (Firm)$_{i,j,t}$	The 10-year rolling average of the firm's headquarters' country inflation rate, obtained from the World Bank.
10-year average real GDP growth (Analyst)$_{i,j,t}$	The 10-year rolling average of the analyst's country real GDP growth rate, obtained from the World Bank.
10-year average real GDP growth (Firm)$_{i,j,t}$	The 10-year rolling average of the firm's headquarters' country real GDP growth rate, obtained from the World Bank.
Default risk (10-year horizon)$_{i,j,t}$	A binary variable equal to 1 if the firm probability of default is above the sample median, and 0 otherwise. The probability of default is measured following the naive Merton's distance-to-default model discussed in Bharath and Shumway (2008), using a horizon of 10 years (T=10).
Default risk (20-year horizon)$_{i,j,t}$	A binary variable equal to 1 if the firm probability of default is above the sample median, and 0 otherwise. The probability of default is measured following the naive Merton's distance-to-default model discussed in Bharath and Shumway (2008), using a horizon of 20 years (T=20).
Disagreement$_{a,b,i,t}$	The difference between analyst A and analyst B TGR estimates squared and scaled by the absolute value of their sum: $(TGR_{a,i,t} - TGR_{b,i,t})^2/\lvert TGR_{a,i,t} + TGR_{b,i,t}\rvert$.
Firm age	The firm post-IPO year age.
$I^{far}_{i,j,t}$	A binary variable equal to 1 if the distance between the analyst's country capital and the firm's country capital is above the sample median, and 0 otherwise. The pairwise distance is measured as the shortest distance between both cities using their centroid GPS coordinates.
LTG (IBES)$_{i,j,t}$	The consensus measure (median) of the LTG variable provided by the Institutional Brokers' Estimate System (IBES), for the firm and the beginning of the calendar year.
Terminal growth rate (TGR)$_{i,j,t}$	The terminal growth rate$_{i,j,t}$ used by equity analysts in their DCF models, measured from the equity reports.
TGR Industry Dispersion	The difference between analysts' terminal growth rate and the average terminal growth rate measured for each age group-industry-year squared and scaled by the industry year average: $(TGR_{i,j,t} - \overline{TGR}^{\text{Age Group}}_{k,t})^2/\overline{TGR}_{k,t}$.

Appendix B Supplementary Figures and Tables

Appendix Figures

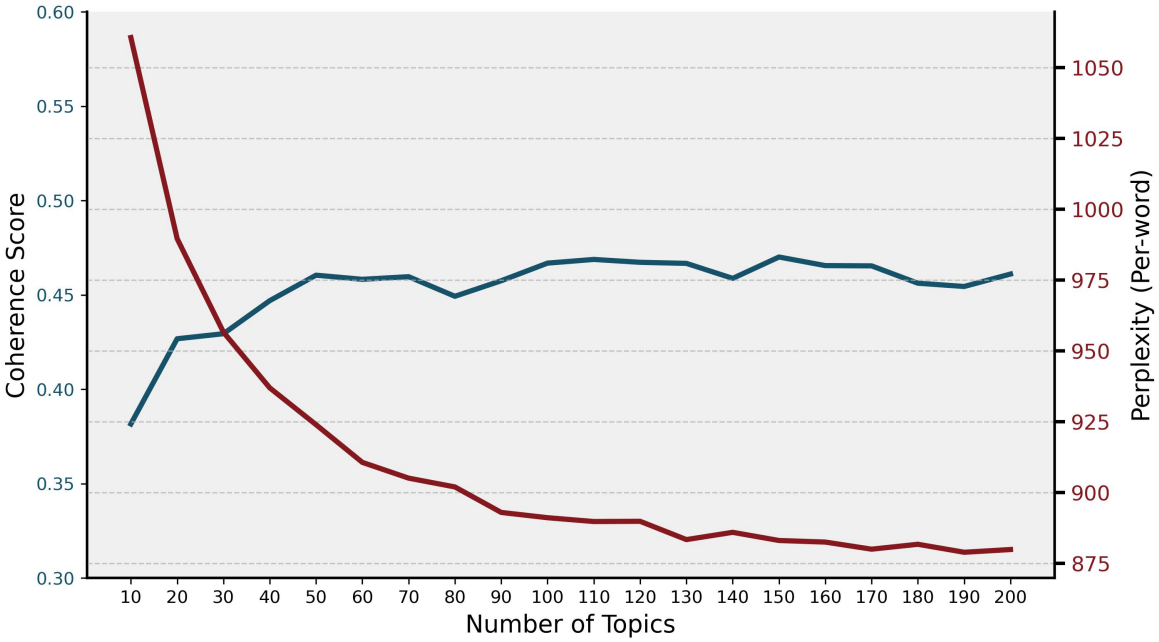


Figure B.1: Coherence and Perplexity Scores This figure plots the coherence and the in-sample perplexity scores when using a different number of topics to estimate the LDA model.

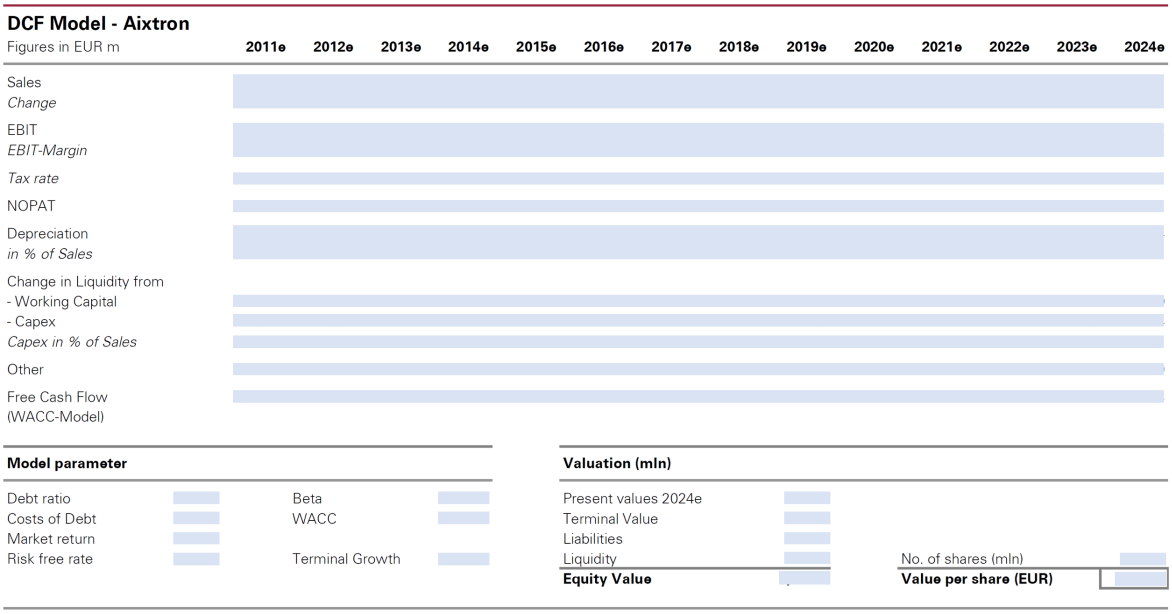


Figure B.2: Example of Complete Equity Report DCF This figure shows a representative example of discounted cash flow models when analysts supplement their recommendations with valuation models. This figure is taken from the Aixtron (Ticker = AIXGn) equity report, published by Warburg Research on October 27, 2011. *For copyright reasons, we redacted any information provided in the table.*

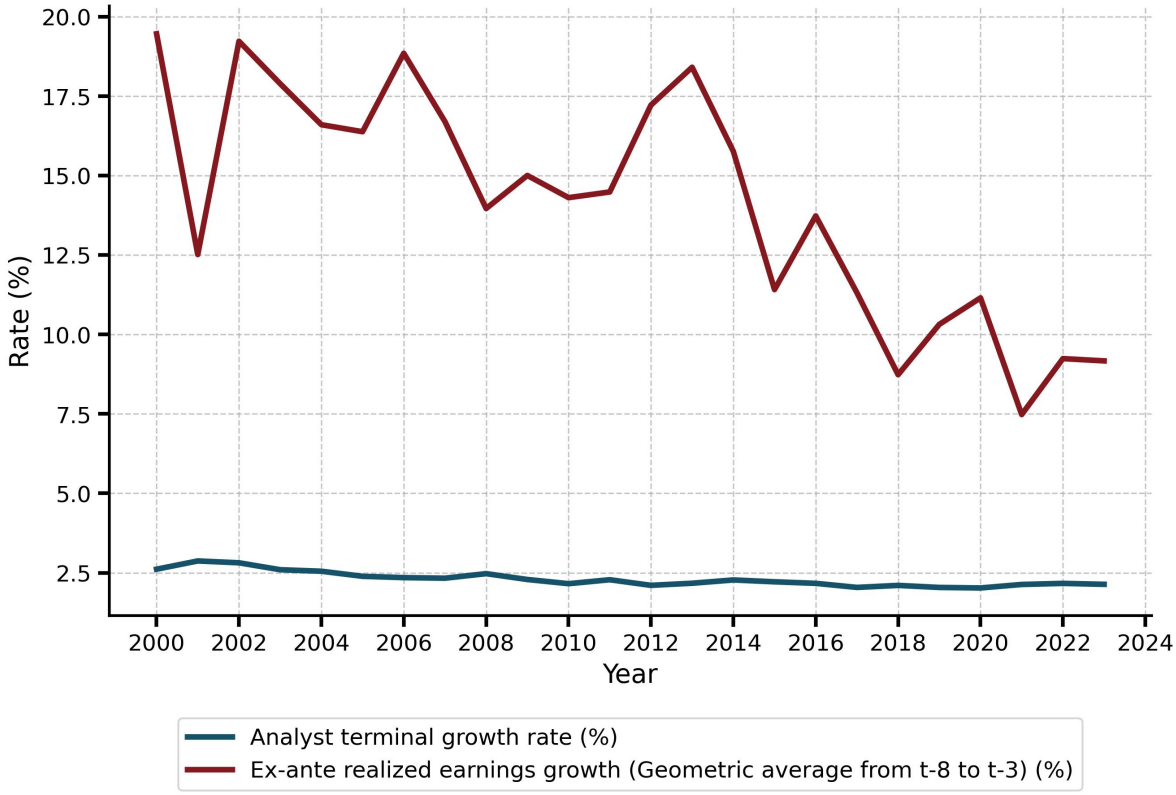


Figure B.3: TGR expectations versus historical returns This figure compares the time trends of historically realized earnings growth at the $t-8$ to $t-3$ horizon with analyst terminal growth rates. Realized earnings growth corresponds to the geometric average realized from $t-8$ to $t-3$, averaged by year (t) in the sample. Terminal growth rate is the average terminal growth rate by year. The x-axis denotes years. The y-axis corresponds to rates, in percentages (%).

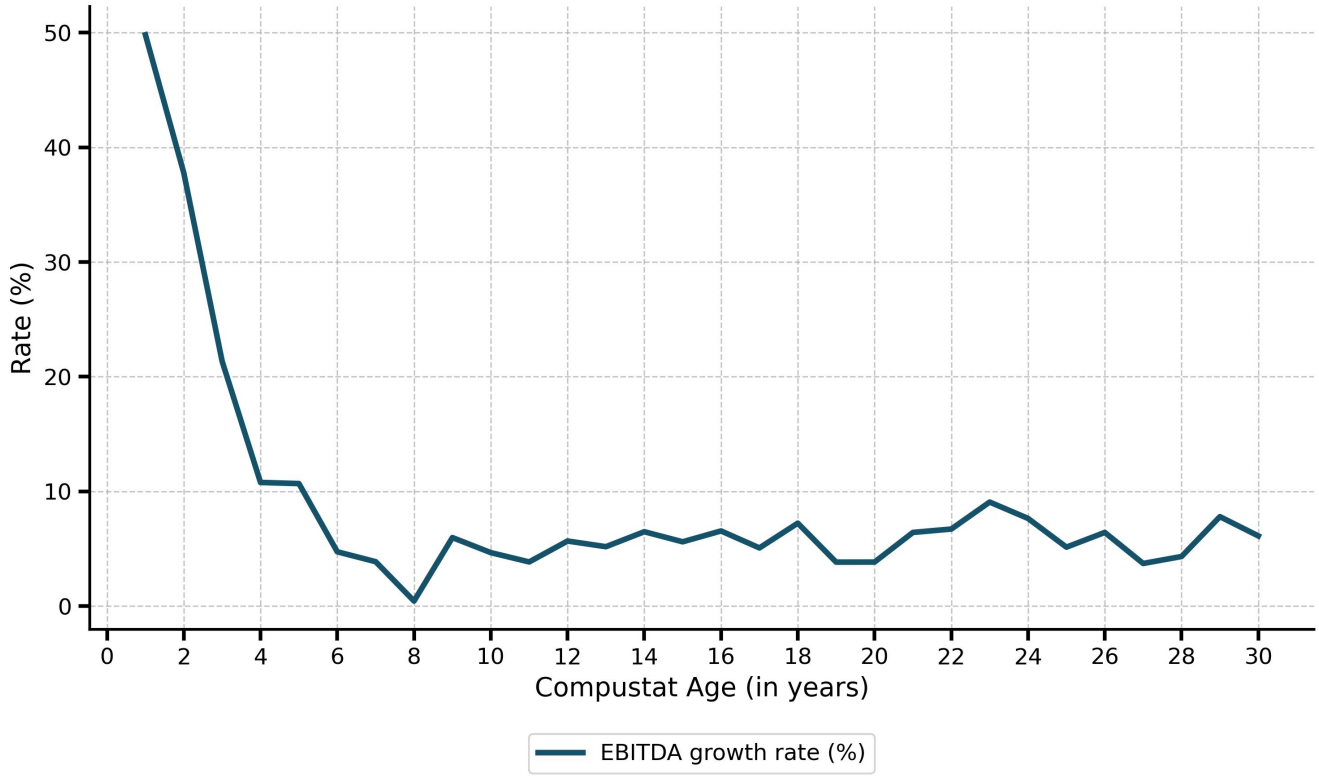


Figure B.4: Earnings Annual Growth Rate and Firm Age This figure plots the realized earnings growth as a function of firm age. The x-axis is the firm age (from the IPO date using Compustat). The y-axis is the average growth rate of earnings. The sample includes all firms included in Compustat for which we have measures of OIBDP (*Operating earnings before depreciation*).

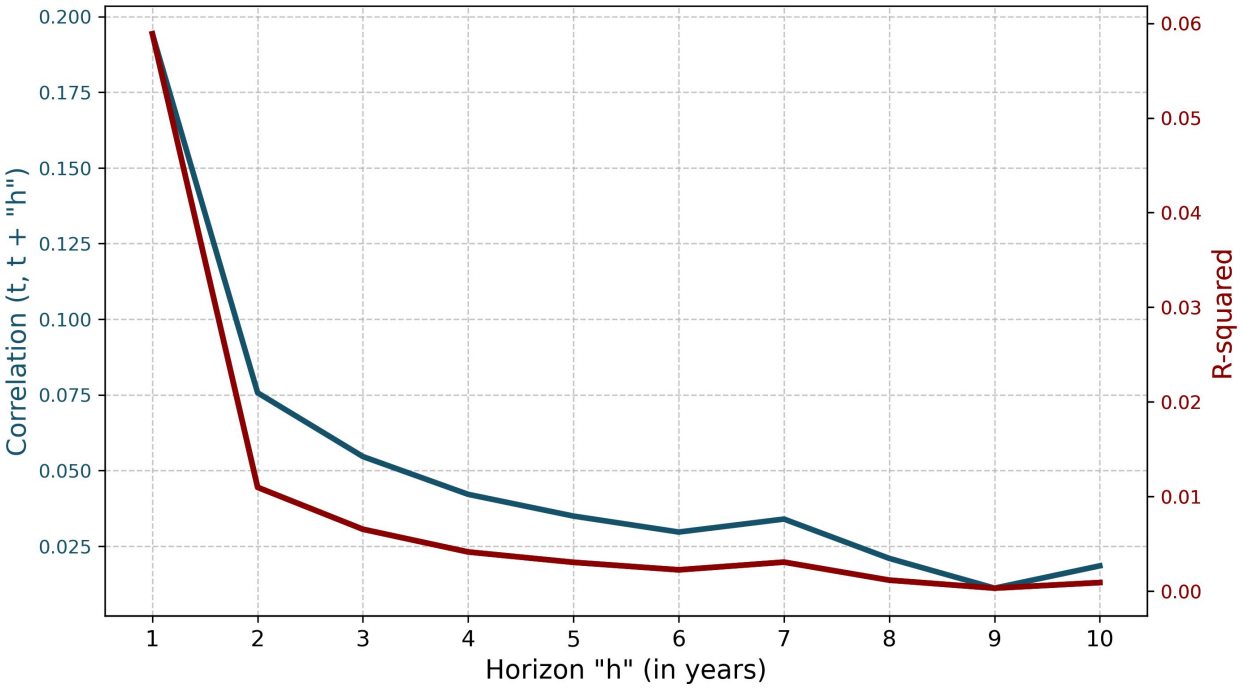
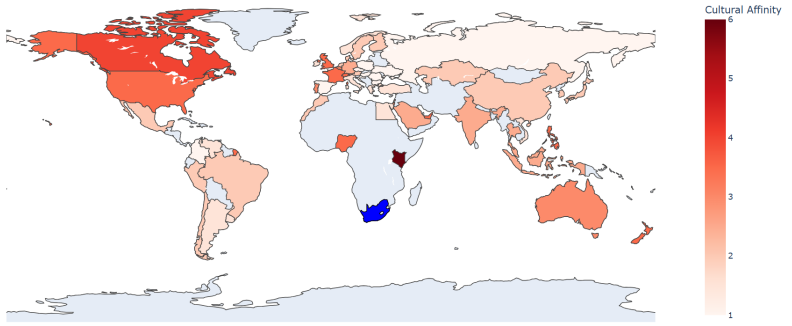
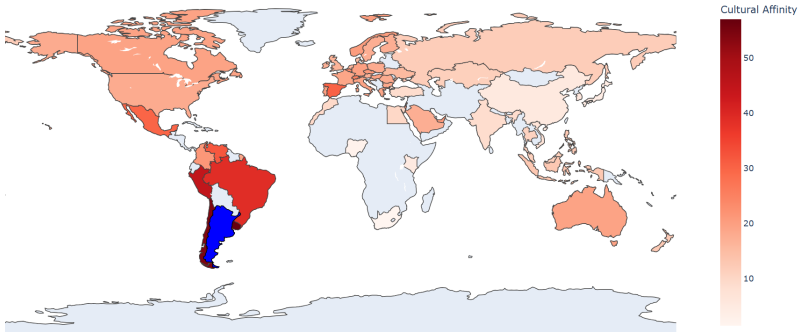


Figure B.5: Correlation realized growth at different horizon This figure plots the average correlation between realized earnings growth rate at time t, and realized earnings growth rate at time $t+h$. In blue, we show the correlate coefficient. In red, we show the r-squared obtain when regression earnings growth at time $t+h$ on earnings growth at time t.

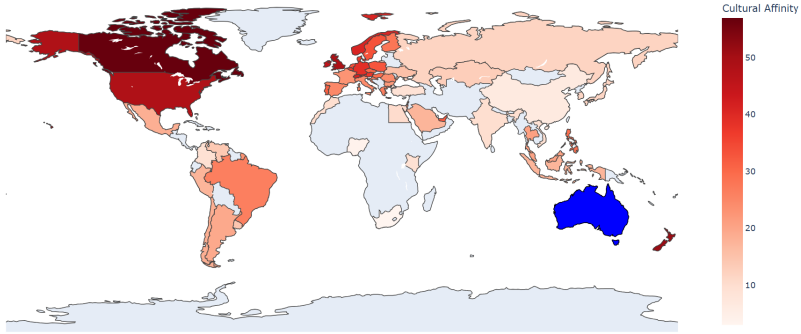
(A) South-Africa.



(B) Argentina.



(C) Australia.



(D) United Arab Emirates.

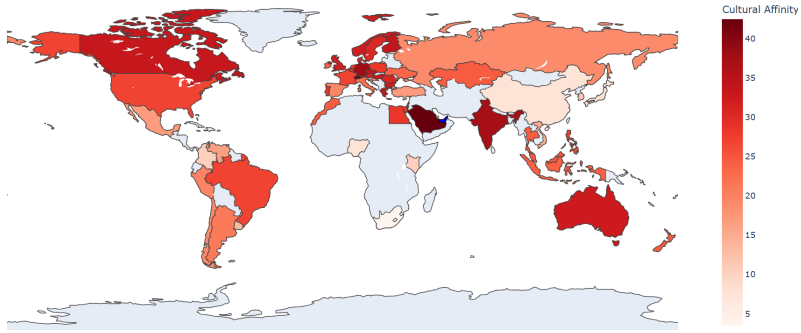
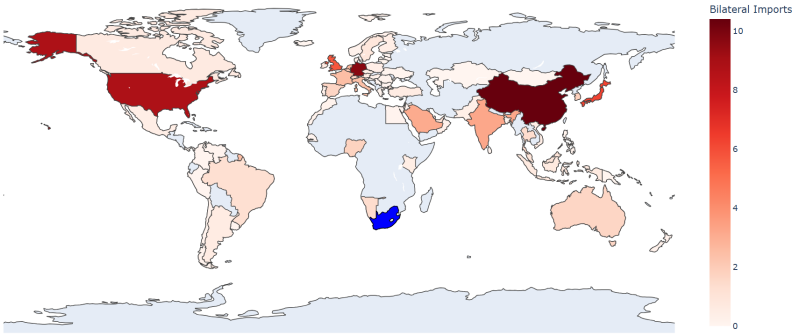


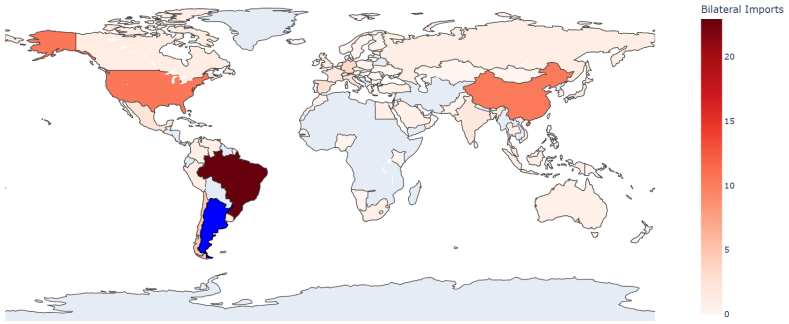
Figure B.6:
Cultural Affinity Between Countries

This figure shows cultural affinity between a focal country and its partner countries, measured by the degree of overlap in their music consumption. Four focal countries are considered: (i) South Africa, (ii) Argentina, (iii) Australia, and (iv) the United Arab Emirates. We measure music consumption overlap using each country's Shazam Top 200 Chart for the week of December 17, 2024. The resulting index is normalized to range from 0 to 100, where 0 indicates no overlap in musical content and 100 indicates a perfect overlap.

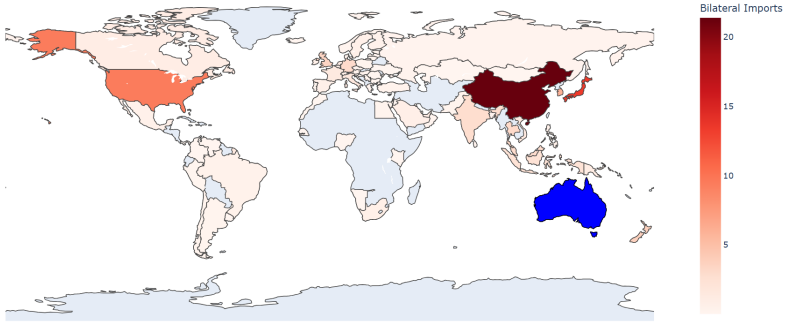
(A) South-Africa.



(B) Argentina.



(C) Australia.



(D) United Arab Emirates.

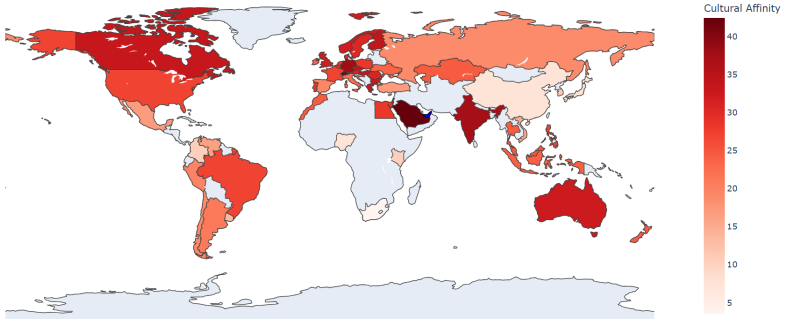
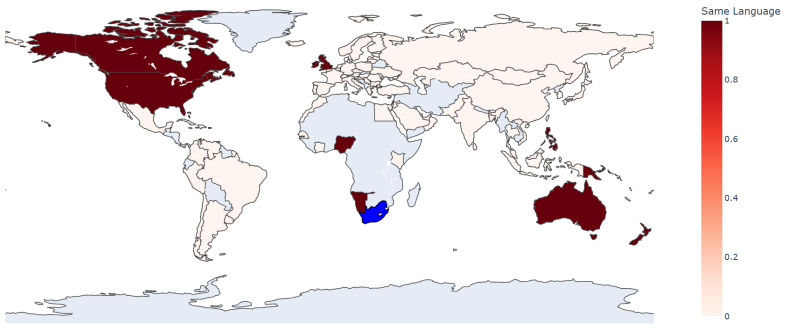


Figure B.7:
Economic ties between countries

This figure shows economic ties between a focal country and its partner countries, represented as the share of total bilateral imports. Four focal countries are considered: (i) South Africa, (ii) Argentina, (iii) Australia, and (iv) the United Arab Emirates. For each focal country–partner country pair, total bilateral imports are defined as the sum of the focal country's imports from and exports to that partner. To assess the relative importance of each trade relationship, we divide the country-pair's total bilateral imports by the focal country's total bilateral imports across all of its partners. The data, sourced from the OECD, cover the period 2000–2021. We take the average share over this entire period for each pair. A value of 0 indicates no trading relationship on average, while a value of 1 indicates that the focal country traded exclusively with that single partner country.

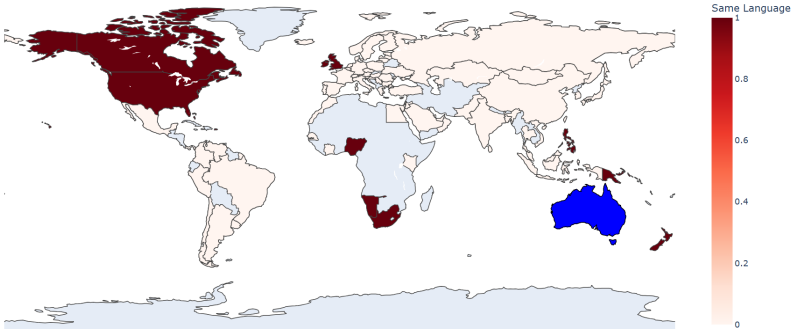
(A) South-Africa.



(B) Argentina.



(C) Australia.



(D) United Arab Emirates.

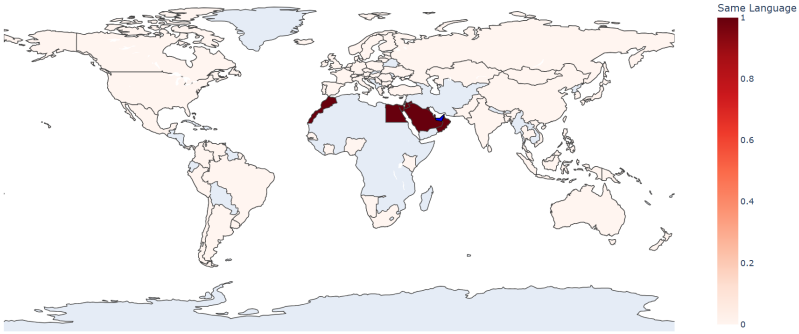


Figure B.8:
Language similarity between countries

This figure shows language similarity between a focal country and its partner countries for four focal countries: (i) South Africa, (ii) Argentina, (iii) Australia, and (iv) the United Arab Emirates. We define a country's language based on its official language. In cases where multiple official languages exist, we select the most spoken one. However, if English is among the official languages, we prioritize English. For countries without an official language (e.g., the United States), we use the most widely spoken language. The language similarity measure takes a value of 1 if the focal country and its partner country share the same defined language, and 0 otherwise.

Appendix Tables

Table B.1: Cosine Similarity Across Different Industries This table reports the Cosine similarity scores for different pairs of firms. The unit of observation is at the analyst a, analyst b, firm i, and forecast year t levels. The sample period is 2000–2023.

	Cosine similarity$_{a,b,i,t}$		
	Different First NAICS Digit (1)	Same First 3 NAICS Digits (2)	Same Firm (3)
Mean	0.74	0.74	0.79
Median	0.77	0.77	0.82
Standard Deviation	0.20	0.19	0.14
25th Percentile	0.60	0.60	0.71
75th Percentile	0.91	0.90	0.89
Number of Observations	277,227,473	6,941,919	193,423

Table B.2: Representative Words from LDA Topics This table reports the main words representing the LDA topics. Words in the list are selected using the reversed frequency weights, where the word weight in a given topic of the LDA model is scaled by its relative frequency from the texts, similar to Bybee et al. (2023b)'s strategy.

Category	Topic Name	Word List
Geography	Asia	rmb, thailand, hong_kong, korea, china, overseas, singapore, india, chinese, domestic
Geography	Europe	spanish, italian, spain, italy, france, french, russian, germany, german, russia
Geography	Oceania	australian, new_zealand, australia, half, earning, forecast, period, earn, result
Geography	South-America	mexican, brazilian, argentina, brazil, chile, mexico, latam, inflation, country
Industry	Health Care	hospital, implant, medicare, pharmacy, diagnostic, medical, care, healthcare, reimbursement, procedure
Industry	Oil and Gas (Transformation)	refining, petrochemical, refinery, feedstock, crude_oil, gasoline, crude, bbl, downstream, oil
Industry	Oil and Gas (Production)	rig, boe, drilling, exploration, field, discovery, reserve, natural_gas, oil_gas, oil
Industry	Manufacturing	equipment, industrial, manufacturing, manufacturer, chemical, system, supplier, component, technology, application
Industry	Entertainment	programming, broadcast, newspaper, television, film, music, radio, audience, entertainment, advertising
Industry	Resources	ethanol, pulp, farmer, sugar, crop, grain, cigarette, corn, seed, fertilizer
Industry	Automobile	battery, dealer, car, truck, vehicle, auto, engine, automotive, man, unit
Industry	Banking	loan, lending, bank, mortgage, banking, insurance, deposit, fund, bond, roe
Industry	Carrier	satellite, tower, fiber, cable, video, carrier, network, voice, speed, broadband
Industry	Communication	arpu, wireline, mobile, subscriber, telecom, churn, broadband, wireless, sub, operator
Industry	Construction	non_residential, toll_road, cement, housing, railway, concession, construction, road, contractor, building
Industry	Consumer goods (Clothing)	apparel, fashion, luxury, brand, wholesale, sport, distributor, channel, inventory, retail
Industry	Consumer goods (Food)	beverage, spirit, tobacco, beer, wine, category, innovation, brand, food, consumer
Industry	Energy	renewable, grid, mwh, hydro, solar, electricity, power_plant, wind, nuclear, utility
Industry	Gaming	gaming, search, game, online, commerce, website, internet, user, platform, marketing
Industry	IT	cloud, software, sap, vendor, solution, enterprise, subscription, application, license, platform
Industry	Mining (Metal Other)	gold, mineral, zinc, south_african, copper, ore, nickel, mining, south_africa, grade
Industry	Mining (Steel)	potash, aluminium, steel, iron_ore, aluminum, coal, tonne, ton, import, producer
Industry	Pharmaceutical (R&D)	generic, pharma, pharmaceutical, patent, drug, fda, royalty, launch, franchise, pipeline
Industry	Pharmaceutical (Clinical trial)	phase_trial, dose, efficacy, therapy, cancer, disease, patient, trial, clinical, study
Industry	Real Estate	reit, tenant, occupancy, property, hotel, rent, rental, real_estate, land, nav
Industry	Retail (Other)	supermarket, store, restaurant, department_store, retailer, format, food, retail, chain, comp
Industry	Semi-Conductor	dram, chip, semiconductor, panel, handset, shipment, tech, device, design, technology
Industry	Shipping	container, vessel, parcel, port, mail, ship, rail, tanker, freight, shipping
Industry	Transportation	airline, aircraft, cruise, passenger, airport, aviation, flight, defense, travel, toll
Macroeconomics	Legal Risk	court, litigation, settlement, legal, card, claim, merchant, affiliate, payment, information
Macroeconomics	Economics Cycles	economy, gdp, index, economic, cycle, recovery, global, weak, sector, decline
Macroeconomics	Regulation	reform, proposal, regulator, regulation, regulatory, government, tax, review, state, policy

Continued on next page.

Continued from previous page.

Category	Topic Name	Word List
Operations	Outsourcing	outsourcing, client, service, contract, employee, provider, revenue, business, segment, acquisition
Operations	Demand	volume, pricing, raw_material, demand, capacity, supply, pressure, increase, recovery, cost
Operations	Expenditure ad Disbursement	dps, dividend, debt, buyback, capex, dividend_yield, cash, fcf, balance_sheet, ebitda
Operations	Management	officer, education, student, school, director, executive, president, esg, cfo, ceo
Operations	Merge and Acquisition	merger, bid, synergy, deal, stake, transaction, shareholder, offer, acquisition
Operations	Organization	disposal, restructuring, division, group, trading, european, profit
Operations	Project Management	waste, project, water, backlog, delay, pipeline, facility, contract, financing, funding
Valuation	Valuation facts	fact, different, reason, important, big, time, long, case, point, issue
Valuation	Accounting Norms and Model Selection	ambiguity_local, modelware_separation, accounting_regulation, value_distortion, financing_complete, fifo_basis, inventory_lifo, eps_adjust, modelware_proprietary, interest_objectivity
Valuation	Free Cash Flow	approximately, roughly, free_cash, flow, facility, cash_flow, company, capital, distribution, acquisition
Valuation	Operating Perfomance	sek, organic, ebit, ebita, currency, sale, europe, margin, effect, growth
Valuation	Profitability	profit, expense, income, loss, operating, gross, charge, net, gain, tax
Valuation	Firm growth	covid, headwind, bps, incremental, leverage, fcf, expansion, organic, upside, margin
Valuation	Recommendation	outperform, neutral, catalyst, recommendation, stock, momentum, buy, rating, earning, earn
Valuation	Guidance	eur, usd, fcf, ebit, slightly, guidance, positive, estimate, good, development
Valuation	Concensus	guidance, street, consensus, estimate, eps, prior, revenue, slightly, expectation, ahead
Valuation	Relative Valuation	multiple, fair_value, historical, peer, analysis, average, discount, relative, scenario, value
Valuation	Risk	beta, wacc, risk, terminal, downside, free, dcf, investment_thesis, hold, unchanged

Table B.3: TGR Disagreement and Textual Narratives (Alternative Topic Model: 150 Topics) This table explores the relation between differences in textual discussion and disagreement in terminal growth rate expectations between pairs of analysts using an alternative version of the LDA specification (150 topics). The unit of observation is at the analyst a, analyst b, firm i, and forecast year t levels. The sample period is 2000–2023. The dependent variable, $Disagreement_{a,b,i,t}$, is equal to the difference between analyst A and analyst B TGR estimates squared and scaled by the absolute value of their sum: $(TGR_{a,i,t} - TGR_{b,i,t})^2/|TGR_{a,i,t} + TGR_{b,i,t}|$. The variables of interest are Jaccard and Cosine scores, which respectively measure the similarity levels between the two analysts' textual valuation theses on the extensive and intensive margins, respectively. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the analyst pair level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable	$Disagreement_{a,b,i,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Jaccard^{150}_{a,b,i,t}$	-0.26	-0.26	-0.37				-0.16
	(0.32)	(0.32)	(0.33)				(0.43)
$Cosine^{150}_{a,b,i,t}$				-0.88**	-0.86***	-0.84**	-0.82**
				(0.35)	(0.33)	(0.34)	(0.32)
Year FE	No	Yes	No	No	Yes	No	No
Industry FE	No	Yes	No	No	Yes	No	No
Industry*Year FE	No	No	Yes	No	No	Yes	Yes
Observations	31,828	31,785	31,765	18,888	18,868	18,841	18,841
F Statistics	0.63	0.67	1.30	6.24	6.84	6.27	3.21
R^2	0.00	0.01	0.05	0.00	0.02	0.07	0.07

Table B.4: TGR Informativeness aggregate This Table replicate the results presented in Table 3 Panel A Columns 3 and 4, our main specification, but using the aggregated measures of Terminal growth rate, and 5-year geometric average of earnings growth. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the analyst pair level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable	Average $(\Pi_{t+4}^{t+8}(1+g_{j,t})^{(1/5)}-1)$	
	(1)	(2)
Average terminal growth rate (%)$_t$	4.24***	5.16***
	(1.11)	(1.56)
Average LTG (%)$_t$		-0.15
		(0.18)
Observations	14	14
F Statistics	14.61	6.65
R^2	0.45	0.47

Table B.5: Trust Toward Other Countries and TGR This table explores the relation between analysts trust toward other cultures and their long-term expectations about firm located in those other countries. The unit of observation is at the firm *i*, analyst *j*, and forecast year *t* levels. The sample period is 2000–2023. The dependent variable, $Trust_{c,k}$, is equal to value of the measure "trust" averaged across all the participant which responded to the survey. The survey data, Eurobarometer survey "ZA Study Number" 2935, was collected in 1997. The sample used in the regression covers all pairs of firm HQ country analyst country that are included in the survey. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable	$TGR_{i,j,t}$	
	(1)	(2)
Trust Index$_{c,k}$	0.41***	0.42***
	(0.11)	(0.15)
Analyst * Year FE	No	Yes
Firm * Year FE	No	Yes
Observations	13,210	3,564
F Statistics	12.87	7.91
R^2	0.00	0.89

Table B.6: OLS alternative approach: Anova variance decomposition This table presents a variance decomposition of analyst terminal growth rates using ANOVA. In Panel A, we replicate the specification used in Column 1 of Panel A in Table 4, and we exclude from the sample firms, analysts, and brokerage houses with fewer than 2 observations in total, as for these observations, their respective fixed effects fully explain the associated variation in terminal growth rates. In Panel B, we exclude from the sample brokerage houses with fewer than 2 observations in total, firms with fewer than 10 observations, and analysts with fewer than 50 observations to replicate the specification used for the results shown in Column 4 of Panel B in Table 4. In both panels, we also exclude brokerage houses that are not associated with equity analysts who switched between employers during the sample period (2000–2023), since it is not possible to distinguish between the effect of those brokerage houses and their analysts (Abowd et al., 1999). Variable definitions appear in Appendix A.

Panel A: Column 1 of Panel A in Table 4	Terminal growth rate$_{i,j,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Brokerage house FE	Yes	No	No	Yes	Yes	No	Yes
Analyst FE	No	Yes	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	No	Yes	Yes	Yes
Observations	36,889	36,701	36,712	36,701	36,712	36,439	36,439
R^2	0.02	0.31	0.33	0.31	0.34	0.53	0.53
Adjusted R^2	0.02	0.24	0.23	0.24	0.24	0.39	0.39
Panel B: Column 4 of Panel B in Table 4	Terminal growth rate$_{i,j,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Brokerage house FE	Yes	No	No	Yes	Yes	No	Yes
Analyst FE	No	Yes	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	No	Yes	Yes	Yes
Observations	5,814	5,814	5,684	5,814	5,684	5,684	5,684
R^2	0.09	0.40	0.51	0.43	0.54	0.60	0.62
Adjusted R^2	0.08	0.39	0.43	0.42	0.46	0.53	0.55

Table B.6: OLS alternative approach: Anova variance decomposition of Analyst FEs This table shows the combined explanatory power of various fixed effects for terminal growth rates and presents a variance decomposition of analyst fixed effects in terminal growth rate expectations. For this analysis, we exclude from the sample firms, analysts, and brokerage houses with fewer than 2 observations in total, as for these observations, their respective fixed effects fully explain the associated variation in terminal growth rates. We also exclude brokerage houses that are not associated with equity analysts that switched between employers during the sample period (2000–2023), since it is not possible to distinguish between the effect of those brokerage houses and their analysts (Abowd et al., 1999). Panel A replicates the specification from Table 7 Panel B column 1. Panel B replicates the specification from Table 7 Panel B column 6. The unit of observation is at firm i, analyst j, and forecast year t levels. Panel B performs a variance decomposition of the analyst fixed effects into 5 dependent variables using OLS. Gender is a binary variable equal to 1 if the analyst is female and 0 otherwise. Race is a categorical variable indicating White, Asian, Hispanic, and other races. Education is a categorical variable indicating master degree, MBA, or other degrees. Birth cohort is a series of birth year indicators. Analysts country is a categorical variable indicating the analyst country at the time of producing the report. Variable definitions appear in Appendix A.

Panel B: Column 1 of Panel B in Table 7	Terminal growth rate$_{i,j,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender FE	Yes	No	No	No	No	Yes	Yes
Education FE	No	Yes	No	No	No	Yes	Yes
Race FE	No	No	Yes	No	No	Yes	Yes
Birth cohort FE	No	No	No	Yes	No	Yes	Yes
Analyst Country FE	No	No	No	No	Yes	No	Yes
Observations	884	884	884	884	884	884	884
R^2	0.00	0.00	0.01	0.09	0.07	0.10	0.16
Adjusted R^2	-0.00	-0.00	0.01	0.04	0.03	0.04	0.07
Panel B: Column 6 of Panel B in Table 7	Terminal growth rate$_{i,j,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender FE	Yes	No	No	No	No	Yes	Yes
Education FE	No	Yes	No	No	No	Yes	Yes
Race FE	No	No	Yes	No	No	Yes	Yes
Birth cohort FE	No	No	No	Yes	No	Yes	Yes
Analyst Country FE	No	No	No	No	Yes	No	Yes
Observations	308	308	308	308	308	308	308
R^2	0.00	0.00	0.00	0.13	0.13	0.15	0.29
Adjusted R^2	-0.00	-0.00	-0.01	-0.00	0.05	-0.01	0.06