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

Paul H. Décaire† and Marius Guenzel‡

April 1, 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, given that firms' current dynamics are often only loosely linked 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 8,010 firms across 49 countries between 2000 and 2023, and 5,246 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, while medium-run expectations do significantly predict shorter-term realized earnings growth (1–5 years, coefficient of 0.32), the estimated relation weakens dramatically at longer horizons (3–13 years, coefficient of 0.04). By contrast, TGR expectations are consistently and strongly associated with 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 (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 indeed negatively correlated with firms' future realized returns, including after controlling for forecasters' *subjective* expected returns in the same valuation report. Controlling for subjective expected returns isolates the portion of belief variation attributable to other factors, particularly subjective growth expectations. Second, we derive that, if contamination were to arise from reverse-engineering growth expectations from prices, subjective expected returns would need to 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 and proxies for time-varying expected returns by investors. 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 associated with 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 on how professionals perform valuation and provide an in-depth analysis of the underlying factors shaping how they 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. 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).

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 with qualitative information—such as text data underlying quantitative forecasts—to shed light on economic agents' belief formation process (e.g., Kendall and

[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 and Myers (2021)) by pointing to variations in the definition of earnings used and the degree of noise introduced by the inclusion or exclusion of special items.

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, 2015), Giglio et al. (2021), De La O and Myers (2021, 2024), Weber et al. (2022), De La O et al. (2024), D'Acunto and Weber (2024), and Cui et al. (2025)). We complement this literature by incorporating self-reported belief data—along with textual explanations of those beliefs—from forecaster-generated documents (equity reports). 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.[7]

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

[7] 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, analysts do 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).

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.

We collect forecasters' terminal growth rate expectations (as well as additional relevant variables such as subjective discount rates) from each report in four steps. First, we pre-process each document using a series of Python programs, including PDF extraction functions and OCR packages, to identify all sections of text, tables, and figures that contain relevant quantitative information for our purposes. Second, we convert these different elements into short text snippets. Third, for each collected variable, we use artificial intelligence, specifically the Open AI API, to extract the numerical value from these snippets.[8] Fourth, while artificial intelligence has proven to be an efficient tool for text extraction Gilardi et al. (2023), and processes complex sentences with a high success rate, concerns may persist that error rates could exceed acceptable levels when left unsupervised. To ensure the highest level of accuracy, we export the text snippets and numerical values extracted by artificial intelligence to Excel and manually verify each extracted value with the help of our research team as an extra layer of quality control.

3.2.2 Analyst Backgrounds from Universe of LinkedIn Data

Next, we merge the TGR dataset on long-run expectations from Section 3.2.1 with the universe of LinkedIn users obtained from Revelio Labs, thereby obtaining additional (self-reported) personal characteristics of forecasters, such as their educational background and (approximate) age. This merge is feasible because the TGR dataset contains key identifiable information on forecasting analysts, including their full name and employer, enabling a clean and unambiguous match with the Revelio data.

In linking the datasets, we require precise name matches between individuals in the TGR and LinkedIn datasets, hand-verifying individuals to account for common name abbreviations. To further ensure the accuracy of our atched dataset, we also require that the LinkedIn-provided employment history aligns with the analyst's employer in the equity reports.[9] Additionally, the name-employer

[8] We conduct steps one to three on the Arizona State University's High-Performance Computing server *Sol*, as these steps are computationally intensive and require a large number of processing nodes along with significant storage capacity.

[9] We account for common variations in employer names (e.g., Bank of America vs. BofA) as well as accented characters (e.g., Societe Generale vs. Société Générale).

match must result in a unique LinkedIn within the dataset. That is, we exclude cases where multiple LinkedIn users with the same name and and the same employer are linked to the TGR dataset.

3.2.3 Summary Statistics

Firms and Coverage (Panel A). Our final sample consists of 41,429 equity reports with extracted information on TGR expectations, from analysts employed by 42 different brokerage houses, covering a total of 8,010 firms. The sample includes expectations data from 5,246 unique forecasting analysts, 2,565 (49%) of whom are successfully matched to LinkedIn. (For forecasters who provide at least five data points in our sample, we take an extra step, including additional manual checking, to link forecasters to LinkedIn, allowing us to match 62% of individual observations in the sample.)

In additional unreported statistics, the average forecasted firm owns assets with a book value of USD 15.8 billion and has an investment rate of 5.55%. These numbers are similar to the IBES universe, where the average firm has assets with a book value of USD 15.2 billion, and an investment rate of 5.3%, suggesting that our sample is comprehensive, representative, and comparable to its commercial counterparts. 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).

TGR Expectations (Panel B). The average subjective TGR belief in the sample is 2.2%, with the 25th-percentile estimate at 1.5% and the 75-th percentile estimate at 3.0%.[10] By contrast, the average IBES LTG consensus forecast is 13.03%, i.e., an order of magnitude larger.[11] We defer a more in-depth discussion contrasting TGR and LTG expectations to Section 4.1.

Analyst and Country Statistics (Panel C). Regarding individual forecaster characteristics, just under 20% of analysts are female. The vast majority of analysts are Caucasian (63%) per the Revelio-based race prediction algorithm. The next largest group is Asian analysts (20%). 48% 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 1971 and the 75th-percentile analyst born in 1985.

The average analyst in the sample contributes 8 reports and TGR forecasts. There are approximately 115 analysts on average who make forecasts for firms located in a given country, and about 140 analysts per brokerage house.

[10] Panel A of Figure 1 plots the average TGR across firms and forecasters 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.

[11] We obtain LTG forecasts from Refinitiv (data item code TR.LTGMedian).

In unreported statistics, 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. Additionally, 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 (also unreported). As a result of the broad geographic coverage in our sample, about 1 in 3 equity reports and TGR forecasts come from foreign forecasters, i.e., analysts who forecast the long-run growth of a firm headquartered in a different country than where they are located. Finally, there are approximately 850 equity reports and TGR forecasts for firms located in a given country.

Economic Data Series (Panels D and E). Panels D and E present descriptive statistics on various economic series for the countries in which forecasted firms and analysts are located. There is meaningful variation in economic conditions of both firm and analyst countries, both in the full sample (Panel D) and when restricting to the subsample consisting of foreign equity reports (Panel E). For example, the interquartile range (IQR) in firm-country real GDP and inflation in Panel D is 2.2pp and 1.9pp, respectively. Similarly, the IQR in analyst-country GDP and inflation in Panel E is 1.7pp and 1.4pp, respectively.

3.3 Methodology: LDA Estimation

After introducing the data and sample, and before examining the relevance and drivers of TGR expectations in Sections 4 and 5, respectively, this section outlines the methodology we use to analyze the relation between quantitative TGR beliefs and the qualitative text discussions in forecasters' reports (Section 5.5). At the core of this exercise, we employ a Latent Dirichlet Allocation (LDA) model to identify the topics discussed in reports and quantity forecasters' attention allocation across topics. We perform the LDA model estimation in two steps: (i) text pre-processing, and (ii) statistical estimation.

Text Pre-Processing. For the topic analysis, we first extract all text (rather than the quantitative information as in Section 3.2.1) from each report using a series of Python routines, now excluding information contained in figures and tables, as well as excluding the legal boilerplate disclaimers at the end of reports. We include a refinement step to address extraction issues that may have been missed by the Python routines, using Meta's LLM, Llama-2-70b.[12] We next apply five filters and transformations to the extracted text data to improve the topic modeling. First, an equity report must contain at least 100 words to be included in the LDA estimation. Second, we remove verbs from the text. Third, we remove stopwords and lemmatize words using standard Python libraries. Fourth, we

[12] For a concrete example, when extracted text lacks spaces between words, LLMs effectively segment sentences into individual words.

organize words into unigrams and bigrams, based on their occurrences in the full set of reports. [13] Finally, for a unigram or bigram to be included in our final text dataset, it must appear in at least 0.1% of all reports used in the analysis.

LDA Model Estimation. Our estimation strategy is similar in spirit to the approach in Bybee et al. (2023a), to which we refer readers for further details on the estimation technicalities. At an intuitive level, the LDA model uses a bag-of-words approach, representing each report in an article-term matrix of dimension $R \times T$. The row indices correspond to the list of R different equity reports, while the column indices (T) correspond to the distinct terms representing the corpus generated through the text pre-processing steps. The numerical values of the individual matrix elements indicate the number of times each term appears in a given 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).

To determine the optimal number of topics (K), we estimate multiple versions of the LDA model, varying the number of topics from 10 to 200 in increments of 10. Our model selection strategy is then based on three metrics: (a) the coherence score, (b) the perplexity score, and (c) a manual review of the keywords within each topic. The coherence score measures model interpretability by assessing the semantic similarity of words within each topic, with a higher score indicating greater similarity. The perplexity score evaluates model fit, where a lower score suggests a better representation of the data. While these metrics provide useful guidance, the manual review of words within each topic ensures their *natural* interpretability. Regarding the remaining two hyper-parameters, we set $\alpha = 0.05$ and $\eta = 0.02$, values chosen based on optimization of the model's perplexity score. These magnitudes are also broadly consistent with standard LDA estimation practices, where a common rule of thumb, such as $\frac{1}{\text{Number of topics}}$, serves as a reasonable starting point for the estimation.[14]

Figure B.1 plots the coherence and perplexity scores across the range of possible topic numbers in the LDA (Appendix Table B.2 reports the keywords included in our preferred model.) As the figure shows, most gains in coherence and in-sample perplexity occur up to 50 topics, after which improvements level off. Manual inspection further confirms that the 50-topic model is parsimonious and contains topics that are meaningful, mostly non-overlapping, and distinctive. We also note that while the coherence score is maximized at 150 topics, our manual inspection reveals that the words in this higher-topic model tend to be more generic, with some overlapping topics. Nonetheless, while our preferred LDA specification models reports using 50 topics, as mentioned, we use the

[13] In this grouping, words that often occur together receive a higher likelihood of being treated as bigrams than two unigrams. For example, when "supply" and "chain" are frequently written consecutively in a text, the likelihood of using the bigram "supply chain," instead of their unigrams, is higher.

[14] We note that our conclusions of the tests using this LDA output are not sensitive to perturbations around those parameter values.

150-topic model to validate the robustness of our results.

4 The Relevance of TGR Expectations

This section establishes the economic significance of analysts' TGR expectations, reflecting their truly long-term cash flow expectations. First, we demonstrate that TGR expectations provide distinct economic insights beyond those captured by other expectation series used in prior research. We show that TGR expectations exhibit limited overlap with the traditional long-term expectation benchmark, IBES LTG. Additionally, TGR expectations are strongly associated with ex-post realized long-term firm earnings growth, whereas IBES LTG—corresponding to analysts' expectations over a 3- to 5-year horizon—better captures medium-term dynamics. Finally, we present novel evidence that analyst expectations—here, TGR expectations—are not contaminated by investors' expected returns.

4.1 Properties of TGR and IBES LTG Expectations

As alluded to in Section 3.2.3, truly long-run TGR expectations and IBES LTG expectations only share limited similarities. As discussed, comparing the levels of both series reveals a systematic and persistent gap, averaging 10.0 percentage points throughout the sample period. This indicates that TGR expectations are substantially lower than IBES LTG expectations (Table 1; see also Panel A of Figure 2). Relatedly, Appendix Figure B.3 shows 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 examined in prior work are systematically above historical growth rates (Nagel and Xu, 2022; Bordalo et al., 2024b).

Additionally, the correlation between TGR beliefs in our sample and IBES' consensus LTG expectation in a given firm-year is, while statistically significant, economically weak ($\rho = 0.07$; see Panel B of Table 1). Panel B of Figure 2 presents a scatter plot illustrating the relationship between TGR beliefs and IBES LTG expectations, further highlighting their weak and scattered association. It is difficult to visually discern a clear relationship between both expectation series, reflecting their considerable dispersion. The figure also clearly shows that IBES LTG exhibits substantially greater volatility than TGR expectations. Indeed, the standard deviation of IBES LTG (22.5 pp) is 17 times greater than that of TGR expectations rate (1.29 pp; see Table 1).

Altogether, the above evidence establishes that TGR expectations differ systematically from other expectation series that capture growth beyond a one- or two-year horizon, such as the IBES LTG series. As a direct implication of this, our understanding of how professional forecasters and market participants form long-term expectations about firms' growth potential remains limited.

4.2 TGR Expectations and Future Realized Firm Earnings Growth

To directly establish TGR beliefs as a meaningful measure of very long-run growth expectations, and further distinguish them from LTG expectations, we next analyze how the two series relate to firms' future (i.e., ex-post) realized long-term growth at various horizons.

Panel C of Figure 1 illustrates the average association between TGR expectations and annualized ex-post realized earnings growth from year $t + 3$ to $t + 8$ over time.[15] The ($t + 3$ to $t + 8$)-window is our preferred horizon given that the modal DCF valuation model—which the TGR expectations in our sample are an input to—sets the explicit forecast horizon to three years and applies a constant terminal growth rate estimate to any subsequent cash flows afterward. Comparing the TGR beliefs with this ex-post measure of realized growth in the figure, we find that both series strongly comove. Both TGR beliefs and ex-post firm growth peaked in the early 2000s and then declined steadily, exhibiting a secular downward trend.

We also note differences in the levels between TGR beliefs and realized growth rates (TGR expectations − ex-post realized growth = −3.2 pp). In light of this, and to formalize the patterns in Panel C of Figure 1, Panel A of Table 2 estimates linear regressions considering three different ways to measure ex-post long-run realized growth: the average geometric earnings growth for each firm in our sample over the periods t to $t + 5$, $t + 3$ to $t + 8$ (as previously), and $t + 3$ to $t + 13$. The last window expands the horizon over which we measure realized growth to a 10-year period, trading off a reduced sample size for a more comprehensive measure of long-run firm growth.[16] Additionally, the formal regression analysis allows us to include firm-country*industry*year fixed effects (in the even columns), as well as the firm-year consensus LTG expectation variable from IBES, thus effectively estimating horse race regressions between TGR and LTG expectations at various horizons.

Focusing first on our preferred specifications in Columns (3) and (4), measuring realized earnings growth over the $t+3$ to $t+8$ horizon, we find that a one percentage point increase in TGR expectations is associated with a 0.46–0.66 percentage point increase in ex-post realized growth, depending on whether we include or exclude fixed effects. This strong, positive relationship between TGR beliefs and realized firm growth persists across the other examined horizons, as shown in Columns (1) to (2) and Columns (5) to (6). For instance, in the specifications with fixed effects, a one percentage point increase in TGR expectations is associated with a 0.55 percentage point increase in realized growth over the next five years, and with a 0.40 percentage point increase over the $t + 3$ to $t + 13$ horizon. The robustness and persistence of the positive relationship between average TGR expectations and

[15] To clarify notation, we use $t+3$ to $t+8$ to refer to the period from the beginning of year $t+3$ through the beginning of year $t + 8$. Similarly, we will later refer to the period from the beginning of year t to the beginning of year $t + 5$ as (t to $t + 5$), and so on.

[16] As Figure B.4 shows, firms' realized earnings growth rates tend to stabilize relatively quickly as they mature. Therefore, examining realized growth rates up to 13 years out captures a significant component of long-run firm growth.

long-run firm growth support the interpretation that TGR estimates are informative about firms' growth trajectory in the (very) long run, establishing their economic substance and significance. Corroborating this conclusion, Table B.4 confirms that TGR expectations account for a significant share of ex-post realized growth in the aggregate, with an R^2 of 0.45.

We next focus on the comparison between TGR and LTG expectations. At IBES' stated forecast horizon for LTG beliefs of 3–5 years (Bordalo et al., 2024b), we find a meaningful economic relationship between consensus LTG expectations and realized firm growth. A one percentage point increase in LTG expectations is associated with a 0.30–0.32 percentage point increase in average realized earnings growth over the $t + 1$ to $t + 5$ forecasting horizon. However, in contrast to TGR expectations, the relationship between LTG and realized firm growth weakens substantially and monotonically as we extend the forecasting horizon. At the $t + 3$ to $t + 8$ horizon, the association declines to just 0.08 percentage points, and at the $t + 3$ to $t + 13$ horizon, it further declines to just 0.04–0.05 percentage points. This pattern highlights the diminishing predictive power of LTG forecasts as the horizon extends, indicating that LTG expectations become increasingly disconnected from realized earnings growth over longer time frames. Consistent with this, in the aggregate time series regressions, LTG beliefs insignificantly predict realized growth, and the R^2 of improves only marginally to 0.47 (compared to 0.45 with only TGR expectations as a regressor; Table B.4).

In sum, the results show that both TGR and LTG are associated with long-run realized earnings growth, but their relationships evolve differently as the forecasting horizon extends. This confirms that the information content of each variable is concentrated on different time horizons of firm growth, with only TGR expectations effectively capturing long-run growth beyond the first few years.[17]

4.3 Reverse-Engineering of TGR Expectations from Prices and Return Contamination

Before we turn to studying the drivers of TGR expectations in Section 5, we first address one potential remaining concern: that analysts may reverse-engineer expectations (i.e., TGR expectations in our case) from prices—a possibility that has recently been the focus of active debate in the literature (cf. Nagel, 2024). Our granular data allow us to make substantive progress on this issue and provide novel evidence against reverse-engineering.

Reverse-Engineering. The core idea behind reverse-engineering is that analysts do not report their own expectations in equity reports, but instead infer investors' beliefs from observed prices. That is, they report high expectations when prices are high, and low expectations when prices are low. Of course, this approach is not inherently problematic, if forecasters accurately extract investors'

[17] This disconnect between shorter-run and truly long-run expectations is mirrored in realized growth patterns, where the correlation between current and historical growth rates steadily declines as the time gap between them increases (Figure B.5).

true long-term beliefs. In that case, it would be equivalent to forecasters independently forming long-term beliefs using a valuation model aligned with that of investors. Under either scenario, extracted TGR estimates would convey meaningful insights into the long-run firm growth beliefs of key financial agents.

The underlying concern, instead, is that investors and forecasters may, unaccounted for by the econometrician, use different valuation models or inputs. Precisely, if forecasters report beliefs (e.g., TGR or other beliefs) inferred from current prices but use discount rates that differ from those used by investors, the resulting reported beliefs may be contaminated by investor returns expectations. In the most extreme case—where forecasters apply a constant discount rate while investors use a time-varying rate with constant growth expectations—a high price today reflects low expected returns, but forecasters will mistakenly infer and report high growth beliefs, and vice versa in the case of a low price today.

Importantly, the most intuitive empirical test for reverse-engineering proves uninformative. Prior literature has documented a negative correlation between longer-term expectations and future realized returns (e.g., Bordalo et al. (2024b)). In fact, we replicate this negative relationship for TGR expectations in Panel A of Table 3, whereas the existing literature has focused on LTG beliefs.[18] While these patterns are consistent with forecasters stating high beliefs when investor expected (and subsequently realized) returns are low, they are *equally* consistent with a world in which forecasters and investors use the same valuation model and inputs, but either overstate or understate firms' growth prospects in different periods—leading to high (or low) prices today and, correspondingly, low (or high) realized returns in the future.

Prior Evidence. Prior work has advanced two lines of argument to suggest that analyst beliefs do not spuriously capture investors' expected returns.

The first shows that analyst expectations continue to predict realized returns even after controlling for proxies of time-varying required returns—narrowing the scope for misattribution or contamination of beliefs by return variation. Recent work by Bordalo et al. (2024b), for instance, finds that LTG expectations retain their negative relationship with realized returns even after controlling for the surplus consumption and consumption-wealth ratios in Campbell and Cochrane (1999) and Lettau and Ludvigson (2001), respectively, as well as the $SVIX^2$ index in Martin (2017).

The second argument provides empirical evidence that investors and forecasters rely on similar frameworks when forming beliefs about asset prices. A large body of work suggests that investors attribute price variation, at least in part, to changes in expected returns and expected earnings growth. Consistent with this, fundamentals such as dividends, earnings, and cash flows help predict

[18] The negative relationship between TGR expectations and realized returns holds both without fixed effects and with year, firm, and firm country*year fixed effects, while controlling for forecasters' subjective cost of equity as a predictor of realized returns (Decaire and Graham, 2024).

returns in market data (Golez, 2014; Sabbatucci, 2022; Hillenbrand and McCarthy, 2024; Pruitt, 2024). Mirroring this, analysts' subjective expected returns and expected earnings growth also inform valuations (Chen et al., 2013; De La O and Myers, 2021; Decaire and Graham, 2024). A wide range of studies further documents that the capital asset pricing model (CAPM) is one of 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). Relatedly, Decaire and Graham (2024) show that forecasters who report TGR estimates within discounted cash flow (DCF) frameworks apply time-varying discount rates—contrasting with the extreme possibility of reverse-engineering above in which analysts use a constant discount rate while investors use a time-varying one with fixed growth expectations. Finally, recent evidence shows that analysts' discount rate estimates track realized future returns on average (Decaire and Graham, 2024), implying that these inputs plausibly reflect the key features of investor return expectations as well.

New Empirical Tests. Building on the combined evidence in prior work, we can formalize the valuation framework used by forecasters in our sample, which in turn allows us to derive new empirical tests for the reverse-engineering concern:

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

Here, p_t^a denotes the analyst's subjective valuation of the firm, and r_t^a is the analysts' subjective, time-varying discount rate. The term $E^* \sum_{j=1}^{N} \rho^j \Delta d_{t+1+j}$ is the analyst's subjective forecast of discrete earnings growth over the next N periods, while from period $N+1$ onward, the forecaster uses their subjective TGR expectation, E^*TGR.[19]

Working from equation 1, the key reverse-engineering concern is that analysts (i) use observed prices to back out beliefs (implying they start from $p_t^a = p_t$, where p_t is the current market price), and (ii) use a *modified* 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 the analyst's and investors' discount rate, and this difference then contaminates the analyst's growth expectations as they reverse-engineer investor beliefs from observed prices. The contamination from ϵ can be distributed across both short- and long-horizon growth expectations, as governed by ϕ, such that equation 2 holds under reverse-engineering.

Two scenarios are possible: (1) all the contamination from using a different discount rate is

[19] That is, equation 1 uses the superscript * on the expectation operator to denote that the beliefs are subjective.

contained in short-term growth expectations ($\phi = 1$), or (2) the contamination affects long-term TGR expectations as well ($\phi \neq 1$). Needless to say, Case (1) does not pose a threat to our analysis of TGR expectations, whereas Case (2) would be problematic.

Examining Case (2) more closely, ϵ implies a direct mechanical relationship between a given analyst's discount rate and their observed TGR estimate. That is, even after controlling for macroeconomic variables likely to affect both r_t^a and TGR beliefs (e.g., GDP, inflation, and interest rates), a residual comovement between the two variables should persist, driven mechanically by ϵ. This yields a new testable implication that we can implement by leveraging the granularity of our data: the absence of comovement between analyst's observed subjective discount rate and TGR beliefs—after controlling for macroeconomic factors likely to jointly drive variation in both variables—would provide evidence against a meaningful role for ϵ, and thus against contamination due to reverse-engineering.

Panel B of Table 3 implements the test. In Columns (1) and (2), we adopt an approach similar to Decaire and Graham (2024), but relate analysts' subjective cost of equity—rather than their subjective discount rates (WACC), as in the prior work—to their subjective TGR expectations. Column (1) includes no fixed effects and finds a small, positive, and statistically significant coefficient. This is consistent with the idea that both TGR expectations and the subjective cost of equity may be associated with common macroeconomic variables (e.g., interest rates). Column (2) introduces firm-country*year fixed effects, which control for macroeconomic conditions, as well as firm fixed effects. In this case, we find an economically small coefficient that is not statistically significant. Both columns yield patterns consistent with those in Decaire and Graham (2024) using subjective discount rates instead.

Finally, in our most restrictive specification in Column (3), we include analyst-firm fixed effects, allowing us to examine how a given analyst updates both inputs for a particular firm over time. The estimated coefficient remains and not statistically significant at conventional levels (p-value > 0.10).[20]

Considering that equation 2 predicts a direct, mechanical—and thus statistically strong—relationship between subjective TGR expectations and expected returns in the presence of return contamination in long-run beliefs, the evidence in Table 3 provides direct evidence against this possibility.[21] Our tests complement and extend the findings in Bordalo et al. (2024b), with our contribution being that we observe all components of forecasters' valuation models and, as a result,

[20] The results in Panel B of Table 3 are robust to expressing TGR and r_t^a in logarithmic form (i.e., $\ln(1+TGR)$ and $\ln(1+r_t^a)$) instead of in percentage terms (i.e., TGR (%) and R_t^a (%), as shown in the table). This rules out both additive and multiplicative analyst errors ϵ.

[21] Instead, the evidence is consistent with two alternative interpretations: (i) analysts engage in reverse-engineering, but only for short-run beliefs, leaving their subjective long-term growth expectations unaffected; or (ii) reverse-engineering is not widespread to begin with and plays a negligible role in shaping analyst expectations.

can—for the first time—analyze the direct relationship between a given analyst's beliefs and expected returns, rather than relying on (possibly noisy) proxy measures from the literature. Overall, we conclude that analysts' long-run TGR expectations are not mechanically inferred from market prices, and next study the determinants of TGR expectations in Section 5.

5 The Drivers of Expectations: TGR vs. Shorter-Term Expectations

Having established that (i) TGR expectations exhibit distinct economic properties compared to other longer-term expectation series studied in prior work, such as IBES LTG (Section 4.1), (ii) TGR expectations correlate with realized firm growth at longer horizons than LTG (Section 4.2), and (iii) TGR expectations are unlikely to be confounded by return contamination due to belief reverse-engineering from prices (Section 4.3), we now investigate the drivers of TGR expectations. Specifically, we conduct a variance decomposition of TGR beliefs, compare their determinants to those of shorter-run expectations, examine how the heterogeneity in TGR beliefs relates to various underlying firm and forecaster characteristics, and "open the black box" behind TGR expectations by linking them to analysts' qualitative text discussions.

5.1 Variance Decomposition

Decomposing TGR Expectations. Table 4 decomposes the variation in TGR expectations across brokerage houses, analysts, and forecasted firms using an ANOVA-based sequential variance decomposition (Columns (1) to (4) of both Panels A and B). Panel A estimates the decomposition on the full sample, excluding firms, analysts, and brokerage houses with fewer than 2 observations, as the respective fixed effects would fully explain the associated variation in terminal growth rate expectations. Panel B restricts the sample to firms with at least ten observations for robustness. In both panels, we present specifications with at least 10, 25, and 50 observations per forecaster, to further address potential concerns about overfitting.

We consistently find that the three sets of fixed effects explain a substantial share— between 60% and slightly above 70%—of the variation in TGR expectations. Brokerage house fixed effects account for only a very small portion of the explained variation, whereas analyst and firm fixed effects drive most of it. For example, in the largest sample in Column (1) of Panel A, brokerage house fixed effects contribute 3% of the explained variation, whereas analyst and firm fixed effects contribute 55% and 41%, respectively. Similarly, in the narrowest sample in Column (4) of Panel B, we estimate contributions of 14%, 55%, and 32%, respectively. That is, analyst fixed effects consistently account for the largest share of the explained variation, suggesting that differences across forecasters play a particularly important role in shaping TGR expectations.

One limitation of the approach in Table 4 is that the ANOVA-based estimation is not fully

indifferent to variations in the sequential order in which the fixed effects are included. As an alternative approach, we estimate standard OLS regressions in which we gradually include the fixed effects one at a time (Table B.6). Panel A and B replicate Column (1) of Panel A and Column (4) of Panel B from Table 4, respectively, representing the two extreme cases of that table. Columns (1) to (3) present the results when including each set of fixed effects individually, effectively estimating an *upper bound* on the amount of variation explained by the inclusion of the given fixed effects. Intuitively, the included fixed effects may be correlated with other (omitted) variables that also influence TGR expectations (see the related discussion in Ayyagari et al., 2008). The remaining columns add various sets of fixed effects jointly.

Using the alternative OLS approach, brokerage fixed effects alone explain at most 2% of the total variation in TGR expectations in the full sample in Panel A, while analyst fixed effects alone explain up to 31%, and firm fixed effects alone account for up to 33%. Additionally, brokerage and firm fixed effects together explain up to 34% of the variation, while adding analyst fixed effects increases the incremental R^2 by 19 percentage points, bringing the total R^2 to 53%. To further ensure that these results are not driven by overfitting due to the high degrees of freedom in the fixed effects, we also examine the adjusted R^2, which penalizes model complexity. The results remain consistent: adding analyst fixed effects to brokerage house and firm fixed effect increases the adjusted R^2 by 15 percentage points, from 24% to 39%. The conclusions from the OLS results for the narrower sample in Panel B are similar.

Decomposing Shorter-Term Expectations. We extend the ANOVA variance decomposition to short-term growth expectations—IBES 1-year EPS growth forecasts—to examine the extent to which the importance of cross-sectional drivers differs between short- and long-run beliefs. The results, presented in Columns (5) to (8) of Panels A and B in Table 4, reveal significant differences. Most strikingly, forecaster fixed effects explain a substantially larger share of the explained variation in TGR beliefs compared to short-run expectations, contributing 22 to 55 percentage points more. Similar to the results for TGR expectations, brokerage house fixed effects explain a negligible portion of the short-run belief 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 short-run beliefs.

Overall, the variance decomposition results underscore the importance of firm factors and individual forecaster characteristics in shaping forecasters' long-run cash flow expectations. These findings motivate a deeper analysis of both firm characteristics and analyst demographics and backgrounds to better understand the large and persistent heterogeneity in TGR expectations across firms and forecasters.

5.2 Firm Characteristics and Theorized Drivers of Long-Run Firm Growth

A natural approach to exploring how TGR expectations relate to firm characteristics is to link cross-firm heterogeneity in TGR expectations to factors associated with the firm life cycle, given that products tend to become more standardized and opportunities for innovation and growth diminish as firms age and industries mature (e.g., Klepper, 1996).

Firm Age. We begin by relating TGR expectations to firm age. As Panel B of Figure 1 shows, forecasters' TGR expectations monotonically decline with firm age. The average TGR estimate in the sample is 2.5% for young firms (aged 10 years and less) and decreases to 0.5% for very old firms, reflecting diminishing growth opportunities as firms mature. Importantly, the magnitudes in the figure reflect the relationship between TGR expectations and firm age cohort *after* controlling for firm and year fixed effects. The OLS regressions in Columns (3) to (5) in Panel B of Table 2 further confirm that the decline in TGR expectations with firm age is also statistically significant. (We discuss the evidence in Panel B of Table 2 in more detail later in this section.)

Industry, Country, and Firm-Specific Drivers. Motivated by the link between TGR expectations and firm age, we next examine when the explanatory power of firm characteristics is most pronounced and how the importance of these features evolves over firm life cycle. In Columns (1) and (2) of Table 5, we find that a firm's industry and headquarters' country explain only a small portion of TGR expectations for young firms. Specifically, these factors explain only 7% of the variation in very young firms aged five years or less. Even 20 to 25 years after a firm's IPO, industry and country explain only about 23% of TGR beliefs. In Columns (3) and (4), 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. Firm fixed effects account for up to 37% of unexplained variation in very recently IPOed firms and up to about 27% for firms aged 20 to 25 years.

These patterns in young firms reverse as firms mature. As Column (2) of Table 5 shows, in old firms, a firm's industry and country explain a significantly larger share of the variation in TGR expectations—49% for 50-year-old firms, and an even greater magnitude for firms aged 70 years and above. Accordingly, for mature firms, the incremental explanatory power of firm fixed effects over industry and country declines significantly.[22]

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, in which forecasters contend with uncertainty about whether

[22] While we report magnitudes for very old firms (55 years and older), we caution that the sample size for these cohorts is smaller—fewer than 100 observations—possibly raising concerns about overfitting. That said, the patterns observed for older firms align with the progression seen among younger cohorts, where sample sizes average several thousand observations per age group.

a new firm will achieve long-term success or face near-termbankruptcy.[23] As forecasters gain information about a firm's long-term viability, expectations for average long-run growth increasingly converge with broader country and industry trends.

As a concrete example of firm-specific factors that is correlated with TGR, Panel B of Table 2 presents evidence on the role long-term default risk. To assess how varying levels of default risk impact the terminal growth rate expectations 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, given our evidence that firm age is an important driver of TGR expectations and is likely correlatee with default probability. Using this specification, we estimate a negative link between TGR expectations and default risk, suggesting that analysts, on average, apply a 0.21 pp discount—a 10% reduction compared to the sample average—to their TGR beliefs about firms with an above-median default probability. This average discount appears reasonable considering that the average probability of default within the high default risk group (above the median) is 5.8%.

Finally, interpreting the downward trend in TGR expectations documented in Panel B of Figure 1 and Panel B of Table 2 through this life-cycle framework suggests that forecasters' favorable beliefs about the growth potential of young firms outweighs potential negative revisions. This pattern aligns with the broader tendency toward (over-)optimism in longer-term forecasts documented in Bordalo et al. (2024b).

Dispersion in TGR Expectations Over Firms' Life Cycle. As a bridging step between analyzing firm and analyst factors and characteristics, we also examine how the within-industry-year *dispersion* of TGR expectations across forecasters varies over the firm life cycle. As Table 6 shows, dispersion in TGR beliefs is highest in young firms. For these firms, the estimates imply that the average dispersion in TGR expectations is about 1 percentage points. As firms mature, the TGR belief dispersion across forecasters substantially declines—by as much as 85% for the most mature firms. These patterns hold across specifications with and without industry*year fixed effects as well when controlling for firm size. Moreover, a Bonferroni-adjusted test for the joint equality of all the firm age coefficients is rejected at the 1% level in all specifications.

Overall, these patterns in TGR belief dispersion suggest that forecasters exhibit greater variation in their access to, processing of, and interpretation of firm-specific information compared to industry and country factors that shape their long-term expectations.

[23] Nearly 25% of public firms ultimately declare bankruptcy, with most failures occurring within 10 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.

5.3 Forecaster Characteristics and Backgrounds

Having established the importance of forecaster fixed effects in Section 5.1 and Table 4, we next examine the extent to which analyst observables help explain the persistent individual heterogeneity, leveraging the availability of forecaster identities and LinkedIn-matched background information in our data.

Panel A of Table 7 reports the total TGR variation explained by firm, brokerage house, and analyst fixed effects, similar to the information in Table 4, across various data cuts (excluding firms, analysts, and brokerage houses with fewer than two observations (left three columns), as for these observations, as their respective fixed effects fully explain the associated variation in TGR expectations; excluding cases with fewer than five observations; analysts with foreign coverage only; and excluding domestic coverage). Panel B presents a sequential ANOVA decomposition of individual analyst fixed effects for each of these subsamples, estimated using a mover design—i.e., tracking analysts who switch brokerage houses within our sample—to separately identify analyst and employer fixed effects, following the labor economics literature (Abowd et al., 1999). Forecaster demographics jointly explain about 30% of the variation in analyst fixed effects—a substantially larger share than found in 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 report that demographics account for at most 10% of belief variation. Demographic factors particularly associated with analyst fixed effects in our sample include birth year (age) and location (country of residence at the time of the report). These findings remain robust across the various estimated specifications.

Several factors may contribute to the greater explanatory power of demographics in forecaster expectations in our setting. First, our panel data includes multiple observations per forecaster, with most analysts forecasting multiple firms over several years, allowing for more precise estimation of forecaster fixed effects. Additionally, our sample is more diverse, including analysts from many different countries, which enables us to disentangle the firm- and analyst-level country effects—unlike prior studies that often rely on more homogeneous samples (e.g., affluent U.S. retail investors affiliated with Vanguard (Giglio et al., 2021)). Finally, the persistence of long-term expectations, which exhibit limited time-series variation and are largely cross-sectional (Decaire and Graham, 2024), likely amplifies the role of individual characteristics relative to the more volatile nature of shorter-term expectations.

To ensure that the conclusions derived from the variance decomposition are robust, beyond their consistency across the various data filters in Table 7, we conduct an additional robustness test. Specifically, to assess whether our results depend on the use of an ANOVA-based variance decomposition, we incrementally add fixed effects to OLS regressions and evaluate the corresponding increases in R^2 (Table B.7). Panels A and B replicate Columns (1) and (6) of Panel B in Table 7,

respectively, representing the two most distinct specifications of the analysis. Columns (1) to (5) in Table B.7 sequentially introduce each fixed effect individually. Examining the R^2 values, birth cohort (R^2 = 0.09–0.13) and analyst geographic location (R^2 = 0.07–0.13) explain a substantial share of the variation in analyst fixed effects, whereas other demographic variables have less explanatory power. Moreover, adding analyst geographic location to a model that already includes gender, education, race, and birth cohort significantly improves explanatory power, with R^2 increasing by 0.03–0.14 percentage points, representing a 60–93% improvement (Columns (6) versus (7)). Overall, these OLS-based results corroborate the ANOVA-based findings in Table 7, with age and location effects emerging as particularly important.

5.4 Forecasters' Personal Economic Environment and Extrapolative Expectations

A natural question is why forecasters with certain backgrounds—such as their geographic exposures—systematically hold higher or lower TGR expectations. This section examines how locally experienced economic conditions shape long-run expectations, including those for geographically distant firms. As a broader takeaway, recalling the relatively limited role of individual forecaster effects in short-term expectations (Table 4), we might expect long-run expectations—being more abstract—to be more malleable to subjective experiences.

Panel A of Table 8 shows how personally and locally experienced economic conditions influence forecasters' long-term firm growth expectations, exploiting the granularity of the data on the geographic locations of both firms and the analysts covering them. In Column (1) and (2), we find that recent GDP growth in the forecaster's country of residence is strongly correlated with their long-term expectations for firms. Notably, this relationship holds even after controlling for the recent economic performance in the firm's country.

In Column (3), we find that the effects of local experiences on TGR expectations are in fact concentrated in forecasts for firms located in geographically distant regions. This pattern is inconsistent with concerns that the relationship is driven by common shocks across neighboring countries. Instead, it supports the interpretation that local extrapolation reflects analysts' heuristic-based responses to higher information-processing costs and limited familiarity with firms in distant locations. To further address potential confounding mechanisms, the regressions include analyst-country*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 local extrapolation is assessed within such dynamics. The inclusion of these fixed effects strengthens the interpretation of our findings as evidence that local economic conditions heuristically shape analysts' beliefs.

Interestingly, in Columns (4) to (6), we find no evidence that local inflation influences long-term expectations when recent inflation experiences are included in the regressions instead of recent local

GDP growth. This result suggests that the extrapolation mechanism operates primarily through macroeconomic productivity factors rather than inflation. This conclusion aligns with prior evidence from Decaire and Graham (2024), which documents no significant relationship between TGR beliefs and inflation.

To further elucidate why and how analysts engage in local extrapolation for distant firms, we next examine three factors related to familiarity and information access in Panel B of Table 8: (i) language similarity, (ii) economic ties, and (iii) cultural affinity. At a high level, we find that local extrapolation is more pronounced when forecasters face greater barriers along these dimensions, further suggesting that heuristic-driven decision-making plays a significant role, particularly in more abstract and distant—geographically, economically, or culturally—contexts.

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, in line with language barriers introducing frictions in the processing of information, such as official disclosures, press releases, or local media, that are more accessible to native speakers. To validate the intuition behind this result, Figure B.6 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 all countries, we observe both proximate and distant regions sharing the same language, ensuring meaningful variation in the data.

Economic Ties. Next, Columns (3) and (4) split the sample at the median level of bilateral trade intensity between the firm's and the analyst's countries.[24] Forecasters are more prone to local extrapolation when economic ties are weak, consistent with the idea that stronger trade relations are associated with greater familiarity with the economic conditions and products of a foreign country. Figure B.7 further corroborates this interpretation by illustrate variation in trade intensity across the same four benchmark countries, where both geographically proximate and distant regions exhibit varying degrees of trade integration.

Cultural Affinity. Finally, Columns (5) and (6) examine the impact of cultural differences, measured using the degree of music consumption overlap between countries.[25] Local extrapolation is greater when cultural overlap per this measure is low, suggesting that familiarity with cultural preferences correlates with the greater integration and proximity to the context of the foreign country. Figure B.8 confirms sufficient variation in cultural exposure across both proximate and distant regions for the same benchmark countries as above.

[24] 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.

[25] 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.

Corroborating the importance of cultural familiarity in shaping expectations, Table B.5 shows that forecasters' TGR expectations are significantly higher for firms located in more trustworthy countries—as measured by *pairwise* trust between the analyst's and the firm's country—using data from the Eurobarometer survey. The relationship between trust and expectations remains highly stable when we include both analyst*year and firm*year fixed effects, thus tightening the estimation by relying on within-analyst-year and within-firm-year variation in trust between the forecaster and the forecasted firm.

Overall, the evidence in Table 8 extends the prior work on local extrapolation (e.g., Kuchler and Zafar, 2019) by documenting strong extrapolative effects across countries and economic environments, even among professional forecasters and for long-run economic variables. Our findings further contribute by highlighting how these effects vary with similarity and familiarity, with extrapolation being particularly pronounced in more abstract and less familiar contexts. These patterns suggest that information acquisition and processing costs, coupled with heuristic-based belief formation, play a central role in shaping extrapolative behavior.

5.5 Opening the Black Box of Quantitative Forecasts: Qualitative Textual Discussions in Analysts' Reports

In the last part of the analysis, we shed light on the mechanisms underlying analysts' quantitative long-run expectations by examining how their long-term expectation formation process is influenced by differential attention to topics relevant to firms' growth prospects. We analyze how analysts allocate attention in the qualitative sections of their reports across topics, considering both the extensive margin—the decision to include a given topic—and the intensive margin—the depth of discussion devoted to covered topics, and explore how variation in topic focus contributes to heterogeneity in analysts' TGR beliefs.

5.5.1 Topical Attention Allocation

We begin with a general exploration of the topics discussed in analysts' equity reports to evaluate the basic properties of topic attention allocation, using the LDA topic model described in Section 3.3. The average report concentrates on a small number topics: specifically, the top seven topics account for 84% of the report's text content, on average. Additional topics receive virtually no weight, with each receiving 0.4% pp of the total attention "budget," on average. Two non-mutually exclusive mechanisms could explain these patterns: (i) a small number of topics may *objectively* capture a firm's growth potential at a given point in time, or (i) analysts may *subjectively* focus on a limited set of topics when forming expectations. We empirically distinguish between these explanations by comparing differences in the topics discussed at the extensive margin by pairs of analysts covering

the same firm at the same point in time.

Extensive Margin Differences. If objective factors primarily drive topic attention, we would expect a substantial 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 they cover. To distinguish between these possibilities, we estimate the Jaccard similarity between the sets of 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 denote the sets of topics discussed in the reports of analysts A and B, respectively; $|A \cap B|$ is the number of topics covered in both reports, and $|A \cup B|$ is the total number of distinct topics discussed in at least one of the reports. Simply put, the Jaccard similarity quantifies the extent of overlap between the two sets of topics, ranging from 0 (no overlap) to 1 (complete overlap). In our baseline analysis, we define a topic as *covered* if it accounts for at least 1% of the qualitative discussion in a report.

For the average analyst pair, 43% of topics covered by at least one analyst are included in both reports, with an interquartile range of 17 pp (33% to 50%).[26] To make this magnitude more concrete, a Jaccard score of 40% could correspond to a case where analyst A covers topics 1, 2, 3, and 4, while analyst B covers topics 1, 2, and 5.

The Jaccard similarity results indicate that analysts tend to agree on a core set of topics that are consistently included in reports. However, substantial differences remain across analysts, indicating the presence of person-specific subjective factors that influence the broader set of each analyst chooses to discuss.

Intensive Margin Differences. In addition to extensive margin differences, we also examine the intensive margin of topic focus—specifically, the weight a forecasters assigns to each topic covered within a 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%). These figures indicate meaningful variation in how forecasters distribute attention across topics within a given report.

To further examine forecaster-specific topic focus at the intensive margin, we assess the similarity in the topics weight across analyst pairs for overlapping topics across their reports. Specifically, we compare the vectors of overlapping topic weights for each pair of analysts covering the same firm at

[26] Importantly, these results are robust to alternative thresholds for defining topic coverage. For example, using a threshold of 5%, instead of 1% yields a Jaccard similarity score of 28%.

the same time, using Cosine similarity as our metric:

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

where **A** and **B** are non-zero vectors representing the weights assigned to overlapping topics by analysts A and B, respectively, and i indexes the overlapping topics. Cosine similarity ranges from -1 (completely opposite weighting) to 1 (identical weighting), with 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 roughly 0.2. While this average similarity across forecasters may appear high, Table B.1 reports the distribution of Cosine similarities in topic weights between firms, depending on whether they operate in the same or different industries, for comparison. Notably, the average Cosine similarity of 0.79 across forecasters is only slightly higher than that observed between firms operating in different industries (0.74). This implies only modest incremental "convergence" in topics weights among analysts forecasting the same firm at the same time, relative to the 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 inclusion) and the intensive margin (topic emphasis).[27]

5.5.2 Differences in Topic Attention and Variation in TGR Expectations

Having established systematic differences in analysts' topic focus, both at the extensive and intensive margin, we now investigate the extent to which these differences help explain the heterogeneity in analysts' long-term TGR expectations. This analysis links the qualitative content of analyst reports to the quantitative variation in beliefs, providing direct insight into the belief formation process.

Specifically, comparing analyst pairs, we ask whether differences in their TGR expectations are driven by variation in the topics they choose to cover, or by differences in how they discuss the same topics. As before, this pairwise approach leverages a distinctive feature of our setting, namely the frequent occurrence of multiple analysts forecasting the same firm at the same time. It also offers a novel strategy for addressing the dimensionality inherent in LDA output by focusing on variation in topics that are not systematically covered across all analysts or firms within a subgroup. This distinguishes the approach from alternative methods in the literature, such as the use of Lasso

[27] We also note that these patterns cannot be driven by differences in report length across analysts. Both the Jaccard and Cosine metrics normalize topic coverage as a proportion of the total report, yielding unit-less measures independent of word counts. To go one step further, when we restrict the sample to report pairs 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%.

regressions (e.g., Bybee et al., 2023a).

We measure disagreement in TGR expectation between analyst pairs as:

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

where A and B denote the two analysts in the pair, i indexes the forecasted firm, and t denotes the forecast year. Intuitively, higher values of $TGR_{A,B,i,t}$ reflect greater relative differences in analysts' TGR expectations.

Panel A of Table 9 presents the results. Columns (1) to (3) report estimates for the extensive margin, using Jaccard similarity as the key variable, while Columns (4) to (6) focus on the intensive margin, using Cosine similarity. Across specifications, the coefficients on both similarity measures are negative, indicating that greater alignment in topic coverage and focus is associated with lower levels of disagreement in analysts' TGR expectations.

We also estimate robust differences in the strength of the association between forecasters' quantitative beliefs and their qualitative topic discussions, depending on whether similarity is measured along the extensive or intensive margin. The extensive-margin Jaccard estimates are, while negative, not statistically significant at conventional levels (p-values ranging from 0.21 to 0.34), indicating that, in fact, we cannot reject a null relationship. By contrast, the intensive-margin Cosine estimates are large (around −0.60) and significant at the 5% level across all specifications. The strong negative link between Cosine similarity and TGR belief agreement obtains in the raw data with no fixed effects (Column (4)), with year and 2-digit NAICS code fixed effects (Column (5)), as well as with year interacted with industry fixed effects. These fixed effects account for broad trends occurring at the industry level over time, and we emphasize that, because both the dependent variable and topic similarity are constructed as differences at the firm-year level, the analysis is inherently conducted within firm-year.

Lastly, Column (7) includes both the Jaccard and Cosine similarity measures simultaneously. The results corroborate the conclusions from the separate regressions in Columns (1) to (6) that intensive-margin differences in topic emphasis are particularly predictive of cross-analyst differences in TGR expectations.

Overall, the findings indicate that variation in very long-term expectations across forecasters is driven more by differences in emphasis on shared topics than by variation in topic selection. More broadly, the findings emphasize that belief heterogeneity can persist even under common information environments, driven by differences in emphasis rather than content.

5.5.3 Understanding Differences in Topic Attention

In a final set of analyses, we examine the determinants and implications of cross-analyst differences in intensive-margin topic focus. We also seek to validate that the results reported in Panel A of Table 9 are driven by *economically intuitive* differences in how forecasters allocate attention across topics.[28]

To this end, we decompose the Cosine similarity measure into five broad categories—(i) macroeconomics, (ii) industry, (iii) geography, (iv) operations, and (v) valuation—based on topics labels we assign through review of the underlying word content of each LDA-derived topic.[29] This categorization allows us to identify which types of content are most strongly associated with differences in long-term expectations.

We then reestimate three variants of the Cosine similarity score, each constructed exclusively belonging to topics from one three categories: macroeconomic, geographic, or industry-related. Panel B of Table 9 reports a negative and statistically significant relationship between each of these category-specific similarity measures and variation in TGR 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 reflected in their larger (absolute) regression coefficients.

Finally, to further examine the role of geographic location, we analyze whether physical distance between forecasters influences how they emphasize topics in their reports at the intensive margin. Columns (1) to (3) of Panel C in Table 9 show that analysts located farther apart are more likely to diverge in how extensively they discuss the same topics. Columns (4) to (6) focus specifically on analysts' emphasis on geographic topics, showing that greater the geographic distance between analysts is associated with greater divergence in how they discuss geographic content. The results imply that, for example, analysts from Argentina and Brazil are more likely to align in their discussions of a firm's geographic footprint than are analysts from Argentina and the U.S. These findings provide additional support for the role of local extrapolation in shaping belief heterogeneity, with variation in topic focus intensity—driven by local conditions and experiences—serving as an underlying mechanism.

6 Conclusion

In this paper, we employ a new, large-scale dataset on terminal growth rate (TGR) expectations for individual firms provided by professional forecasters. Our data includes not only analysts' long-run

[28] While, as discussed, the LDA model is well suited for our purposes of identifying thematic structures in reports, the topics it generates do not carry inherent semantic meaning. Verifying the content of these topics is therefore essential to ensure that the observed variation in topic emphasis is not merely an artifact of the unsupervised modeling process.

[29] Appendix Table B.2 provides the complete list of LDA-derived topics and their assigned categories.

quantitative forecasts but also their accompanying qualitative discussions, analyst identities, and individual backgrounds, enabling us to study the origins and drivers of very long-run cash flow expectations.

We establish a series of novel findings. First, TGR expectations provide distinct economic insights relative to more commonly studied 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 performance. Second, long-run expectations are not contaminated by expected returns due to reverse-engineering of beliefs from prices. Third, there is persistent heterogeneity in TGR expectations across forecasters—greater than the individual-level variation in short-term expectations. This heterogeneity is especially pronounced for less mature firms where country and industry anchors are weaker and forecaster demographics account for a sizable share of the variation. Fourth, forecasters systematically extrapolate from the economic conditions of their country of residence when forming long-run expectations about geographically distant firms in geographically distant countries—a pattern that is especially pronounced when firms are geographically and culturally distant, consistent with the use of belief heuristics in the absence of familiar or easily interpretable local information. Finally, differences in the thematic focus of forecasters' qualitative discussions emerge as an important mechanism underpinning both long-run belief heterogeneity and extrapolation behavior.

Our findings offer new guidance for macro-finance models and models of belief formation. In light of the limited evidence on truly long-term beliefs, our results provide rare empirical insight into how individuals form expectations over distant horizons. A central feature that emerges is the substantial heterogeneity in long-run beliefs—heterogeneity that may be magnified precisely because long-run forecasts cannot be quickly validated against realized outcomes, making them more abstract and susceptible to personal experiences. A second key ingredient is heuristics-based extrapolation, especially from salient local information and in less familiar contexts. A further takeaway is the close connection between forecasters' quantitative beliefs and their qualitative discussions of beliefs. Overall, our findings invite further research into the long-run belief formation mechanisms we identify, as a path toward improving our understanding of price dynamics and capital allocation—especially given that long-run expectations anchor the bulk of asset valuations.

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.

Campbell, J. Y., and J. H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107:205–251.

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.

Coibion, O., and Y. Gorodnichenko. 2015. Information Rigidity and the Expectations Formation Process: A Simple Framework and New Fact. *American Economic Review* 105:2644–2678.

Cui, T., R. Delao, and S. Myers. 2025. The Subjective Belief Factor. *Working Paper* .

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* .

De La O, R., and S. Myers. 2024. Which Subjective Expectations Explain Asset Prices? *The Review of Financial Studies* 37:1929–1978.

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.

Lettau, M., and S. Ludvigson. 2001. Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance* 56:815–849.

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.

Martin, I. 2017. What is the Expected Return on the Market? *The Quarterly Journal of Economics* 132:367–433.

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).

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.

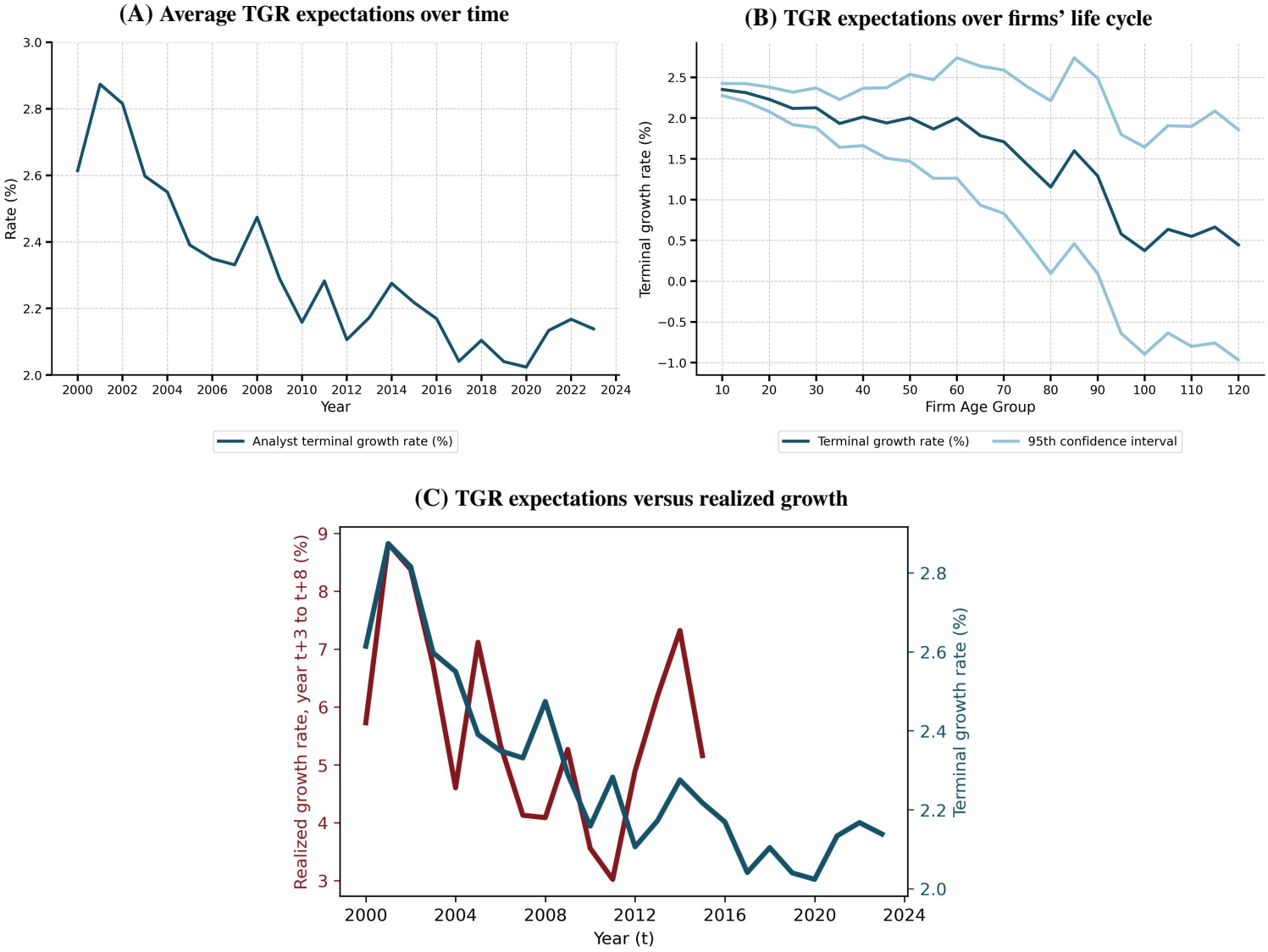


Figure 1: Properties and drivers of terminal growth rate (TGR) expectations Panel A plots the time trend of average terminal growth rate (TGR) expectations across analysts and firms over the sample period from 2000 to 2023. Panel B plots TGR expectations over firms' life cycle. The solid blue line corresponds to age-group fixed effects obtained from the regression: $TGR_{i,j,t} = k + I_{\text{Age group}} + I_{\text{Firm}} + I_{\text{Year}} + \epsilon_{i,j,t}$, where we add back the regression constant (k) to better illustrate the terminal growth rate level. The pale blue lines correspond to 95^{th} confidence interval of the $I_{\text{Age group}}$ regression estimate, where we also add back the regression constant (k). Firm age-group captures firm age groups measured in 5-year increments such that, e.g., firm age-group = 10 corresponds to firms between 5 and 10 years old. Panel C compares the time trends of ex-post realized earnings growth at the $t+3$ to $t+8$ horizon with forecasters' average ex-ante terminal growth rate expectations. 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 expectation by year. In Panel C, the x-axis denotes years, the left-hand-side y-axis shows ex-post realized returns, and the right-hand-side y-axis displays terminal growth rate expectations.

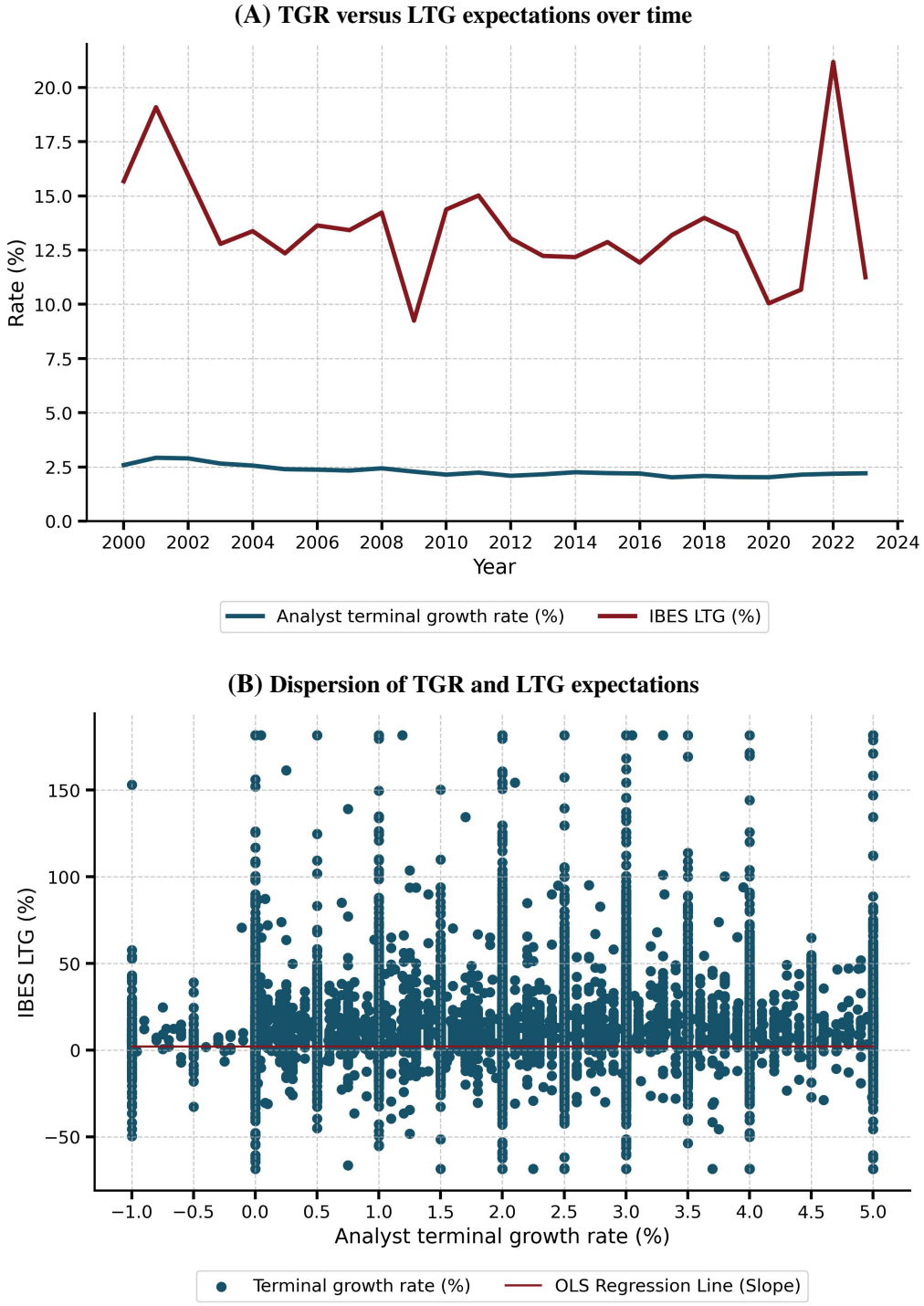


Figure 2: TGR versus IBES LTG expectations Panel A compares the levels and time trends of the TGR expectations in our sample and IBES' LTG expectations. The x-axis shows years and the y-axis is expressed in percent. Panel B shows the dispersion in TGR and LTG expectations. The x-axis represents analysts' TGR expectations in our sample, and the y-axis corresponds to the firm-year LTG consensus (median) in IBES (winsorized at the 0.1 and 99.9th percentiles), with each blue dot representing one TGR-LTG-consensus pair. The red line is the slope from an OLS regression of IBES firm-year LTG consensus beliefs on the TGR expectations in our sample.

Table 1: Summary statistics This table reports summary statistics. The sample includes 41,429 equity reports and terminal growth rate (TGR) expectations for 8,010 firms, obtained from equity reports published between 2000 and 2023. Panel A describes the sample coverage. Panel B reports statistics on the TGR expectations. Panel C presents demographic and other statistics on the forecasting analysts as well as country statistics. Panels D and E present various economic data series. Variables definitions appear in Appendix A.

Panel A: Firms and Coverage	N					
No. Observations (TGR Expectations Sample)	41,429					
No. Firms	8,010					
No. Analysts	5,246					
No. Analysts (LinkedIn-matched)	2,565					
No. Brokerage houses	42					
Panel B: TGR Expectations	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	N
Terminal growth rate (TGR) Expectation $_{i,j,t}$(%)	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 and Country Statistics	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	N
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. Analysts 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
1(Foreign report)	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
Panel D: Economic Data Series—Full Sample	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	N
Firm country						
Real GDP $growth_{i,t}$ (%)	2.55	1.55	2.29	3.74	3.36	38,370
$Inflation_{i,t}$ (%)	2.57	1.26	2.01	3.16	2.50	38,247
10-year Treasury $yield_{i,t}$ (%)	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.	N
Analyst country						
Real GDP $growth_{j,t}$ (%)	2.01	1.45	2.29	3.12	3.70	12,296
$Inflation_{j,t}$ (%)	2.49	1.45	2.27	2.85	1.93	12,285
10-year Treasury $yield_{j,t}$ (%)	3.02	1.51	2.92	4.14	2.10	12,828
Firm country						
Real GDP $growth_{i,t}$ (%)	2.11	0.96	2.21	3.82	3.58	11,802
$Inflation_{i,t}$ (%)	2.43	0.90	1.83	3.06	2.75	11,686
10-year Treasury $yield_{i,t}$ (%)	3.02	0.97	2.75	4.11	2.76	11,830

Table 2: Economic relevance of TGR expectations Panel A regresses ex-post realized firm earnings growth at different horizons on analysts' terminal growth rate expectations and IBES consensus LTG variable. The unit of observation is at the firm i, analyst j, and forecast year t level. The sample period is 2000–2023. The dependent variable is the geometric average of the annual earnings growth rates experienced by firm i over five- or ten-year periods at various horizons (t to $t+5$, $t+3$ to $t+8$, and $t+3$ to $t+13$, respectively). *Terminal growth rate*$_{i,j,t}$ is the analyst's terminal growth rate expectation, measured in percent. $LTG_{i,t}$ is the IBES consensus LTG belief in percent at time t for firm i. Panel B studies the relationship between TGR expectations and default risk, measured using the Merton Model adjustment from Bharath and Shumway (2008), as well as firm age. 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	$\Pi_{t}^{t+5}(1+g_{i,t})^{1/5}-1$		$\Pi_{t+3}^{t+8}(1+g_{i,t})^{1/5}-1$		$\Pi_{t+3}^{t+13}(1+g_{i,t})^{1/10}-1$	
	(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 Consensus LTG (%)$_{i,t}$	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)$_{i,t}$	-0.19***			-0.25***	
	(0.04)			(0.07)	
Default risk (20-year horizon)$_{i,t}$		-0.20***			-0.26***
		(0.03)			(0.06)
ln(Firm post-IPO age)$_{i,t}$			-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 3: TGR expectations, cost of equity, and realized returns Panel A studies the relationship between realized 1-year total equity returns, TGR expectations, and subjective cost of equity estimates. Realized total equity returns are taken from Refinitiv Eikon. Subjective cost of equity is obtained by computing $r_f + \beta * ERP$, where all three subjective inputs are taken directly from the equity reports. Panel B studies the relationship between TGR expectations and the subjective cost of equity estimates. 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 (%)$_{i,t+1}$**		
	(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 4: Variance decomposition of TGR and short-run expectations (ANOVA) This table presents a variance decomposition of analyst TGR and EPS growth expectations using ANOVA. We exclude firms, analysts, and brokerage houses with fewer than two observations, as their respective fixed effects fully explain the associated variation in TGR expectations. We also exclude brokerage houses that are not associated with equity analysts who switched employers during the sample period (2000–2023), as it is not possible to separate the effects of these brokerage houses from those of their analysts (Abowd et al., 1999). Panel A includes the full subsample meeting these criteria (with additional sample filters on the number of reports per analyst specified in the column headers), while Panel B further excludes firms with fewer than ten observations. 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) to (4) perform the variance decomposition for the TGR expectations, while Columns (5) to (8) apply it to the IBES 1-Year EPS growth forecasts. The IBES sample, covering the period 2000–2023, includes only firms also present in the TGR sample to allow for a direct comparison. The IBES 1-Year EPS growth forecast is based on the earliest forecast made by an analysts for a given fiscal year, provided that the forecast is issued at least nine months and at most one year before the target date. Analyst expected growth is measured as the EPS forecast divided by the most recent realized EPS of the firm at the time of the forecast, using IBES (ACTUAL) data, minus one. Variable definitions appear in Appendix A.

Panel A: Anova

	TGR				1-Year Expected EPS Growth Rate			
	Full sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$	Full 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			
	Full sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$	Full 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 analyzes how firm life cycle factors contribute to explaining TGR expectations. The unit of observation is at the firm i, analyst j, and forecast year t level. The sample period is 2000–2023. Column (1) reports the R^2 obtained from estimating the Case 1 Stage 2 regression displayed below for firms within the age range specified in the first column. Column (2) presents the share of the Case 1 Stage 2 regression R^2 attributable to *Firm country* and *Industry* factors, derived through a sequential ANOVA decomposition of the Case 1 Stage 2 specification. Column (3) reports the R^2 obtained from estimating the Case 2 Stage 2 regression for firms within the age range specified in the first column. Column (4) provides a conservative back-of-the-envelope calculation to estimate the share of TGR expectations in Case 1 Stage 1 explained by the firm factors, after accounting for firm country and industry. This estimate is calculated by computing $(1 - R^2_{Column\ 1}) \times R^2_{Columne\ 3}$. 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**	**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: Dispersion in TGR expectations This table presents the results of the dispersion of TGR expectations over firms' life cycle. The unit of observation is at the firm i, analyst j, and forecast year t level. The sample period is 2000–2023. The dependent variable is the squared difference between an analyst's TGR belief and the industry (k) consensus for the firm's age group in forecast year t, scaled by the industry average in that year. The independent variables are indicator variables equal to one if the firm's age falls within the specified range and zero 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%.

Dispersion in TGR Expectations	$(TGR_{i,j,t} - \overline{TGR}_{k,t}^{\text{Age Group}})^2/\overline{TGR}_{k,t}$		
	(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 $\text{Size}_{i,t}$			0.03***
			(0.01)
$\text{TGR}_{i,j,t}$			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 equal (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 presents the combined explanatory power of various fixed effects for TGR expectations in Panel A and presents a variance decomposition of analyst fixed effects in TGR expectations in Panel B. We exclude firms, analysts, and brokerage houses with fewer than two observations (left three columns), as for these observations, as their respective fixed effects fully explain the associated variation in TGR expectations. We also present robustness excluding observations with fewer than five observations (right three columns). We also exclude brokerage houses that are not associated with equity analysts that switched between employers during the sample period (2000–2023), as it is possible not separate the effects of these brokerage houses and from those of their analysts (Abowd et al., 1999). Columns (1) and (4) use the entire respective subsample, Columns (2) and (5) restrict the analysis to firms with both domestic and foreign coverage, and Columns (3) and (6) focus on the foreign coverage of firms included in Column (2) or (5), respectively. Panel A shows the combined explanatory power firm, brokerage house, and analyst fixed effects, with the unit of observation at firm i, analyst j, and forecast year t level. Panel B conducts a variance decomposition of analyst fixed effects into five categories using ANOVA: gender(binary variable equal to 1 if the analyst is female and 0 otherwise), race (categorical variable indicating White, Asian, Hispanic, and other races), education (binary variable equal to 1 if the analyst has a graduate degree and 0 otherwise), and birthyear (birth year indicators). 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. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Panel A: Fixed effects	**Terminal growth rate**$_{i,j,t}$					
Sample filters	At least 2 observations per firm, brokerage house, and analyst			At least 5 observations per firm, brokerage house, and analyst		
	Full Sample (1)	Has foreign coverage (2)	Exclude domestic coverage (3)	Full Sample (4)	Has foreign coverage (5)	Exclude domestic coverage (6)
Constant	2.17*** (0.00)	2.07*** (0.00)	2.11*** (0.00)	2.14*** (0.00)	2.06*** (0.00)	2.09*** (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, brokerage house, and analyst			At least 5 observations per firm, brokerage house, and analyst		
	Full Sample (1)	Has foreign coverage (2)	Exclude domestic coverage (3)	Full Sample (4)	Has foreign coverage (5)	Exclude domestic coverage (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, local economic conditions, and extrapolation This table examines the effect of analysts' local economic conditions on their TGR expectations. The unit of observation is at the firm i, analyst j, and forecast year t level. 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 percent. The 10-year average real GDP growth rates for both the firm's and the analyst's country are obtained from the World Bank and calculated as 10-year trailing averages. Similarly, the 10-year average inflation rates for both the firm's and the analyst's country are constructed using World Bank. as 10-year trailing averages. I^{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. Distance between countries is measured as the shortest distance between their 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)$_{i,t}$	0.13***	0.09***	0.09***			
	(0.02)	(0.03)	(0.03)			
10-year avg. real GDP growth (Analyst)$_{j,t}$		0.09***	-0.04			
		(0.03)	(0.05)			
10-year avg. real GDP growth (Analyst)$_{j,t} * I^{far}_{i,j,t}$			0.17***			
			(0.06)			
10-year avg. inflation rate (Firm)$_{i,t}$				0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
10-year avg. inflation rate (Analyst)$_{j,t}$					0.00	-0.00
					(0.01)	(0.00)
10-year avg. inflation rate (Analyst)$_{j,t} * I^{far}_{i,j,t}$						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)$_{j,t} * I^{far}_{i,j,t}$	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)$_{i,t}$, 10-year avg. real GDP growth (Analyst)$_{j,t}$					
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 examines the relationship between differences in textual discussions in equity reports and disagreement in TGR expectations between pairs of analysts. The unit of observation is at the analyst a, analyst b, firm i, and forecast year t level. The sample period is 2000–2023. The dependent variable in Panels A and B, $Disagreement_{a,b,i,t}$, is defined as the squared difference between analyst a's and analyst b's TGR beliefs, scaled by the absolute value of their sum: $(TGR_{a,i,t} - TGR_{b,i,t})^2/(|TGR_{a,i,t} + TGR_{b,i,t}|/2)$. The key independent variables are the Jaccard and Cosine scores, which measure the similarity levels between the two analysts' textual discussions along the extensive and intensive margins, respectively. Panel C relates the Cosine similarity scores to analyst distance. 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 $^{\text{Macro}}_{a,b,i,t}$	-1.32**	-1.29**					
	(0.61)	(0.60)					
Cosine $^{\text{Industry}}_{a,b,i,t}$			-0.97**	-0.90*			
			(0.46)	(0.48)			
Cosine $^{\text{Geography}}_{a,b,i,t}$					-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 $^{\text{Geography}}_{a,b,i,t}$			
	(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, and j indicates a forecaster.

Variable	Definition
10-year average inflation rate (Analyst)$_{j,t}$	The 10-year rolling average of the analyst's country inflation rate, obtained from the World Bank.
10-year average inflation rate (Firm)$_{i,t}$	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)$_{j,t}$	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)$_{i,t}$	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)$_{i,t}$	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)$_{i,t}$	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$_{i,t}$	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)$_{i,t}$	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

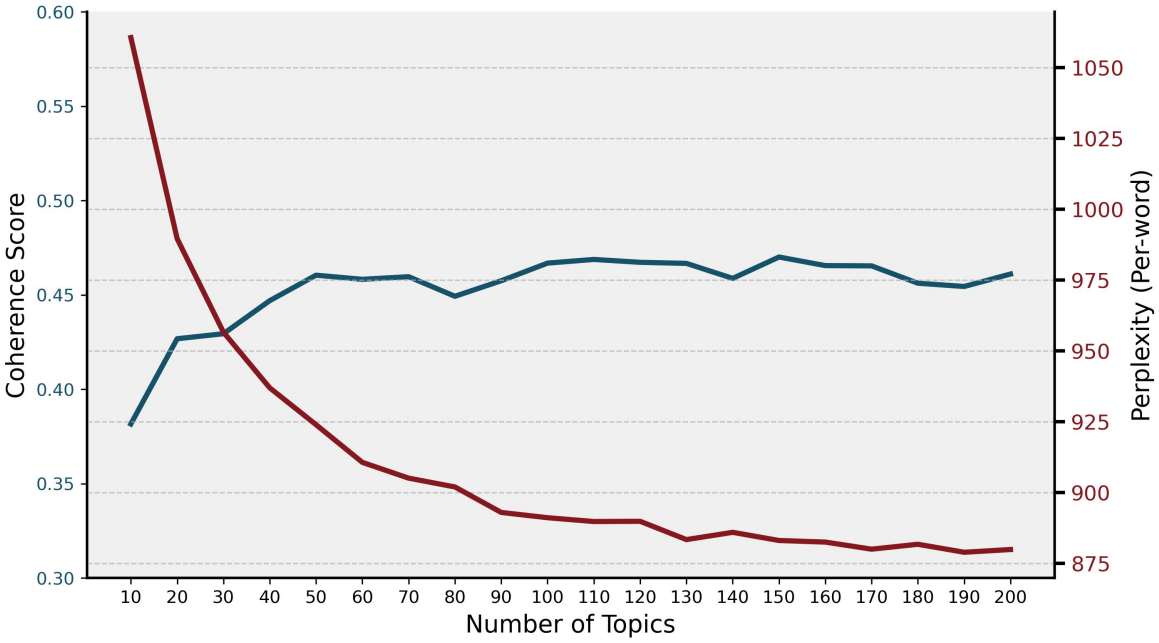


Figure B.1: Coherence and perplexity scores This figure plots the coherence and in-sample perplexity scores for different numbers of topics used in the LDA model estimation.

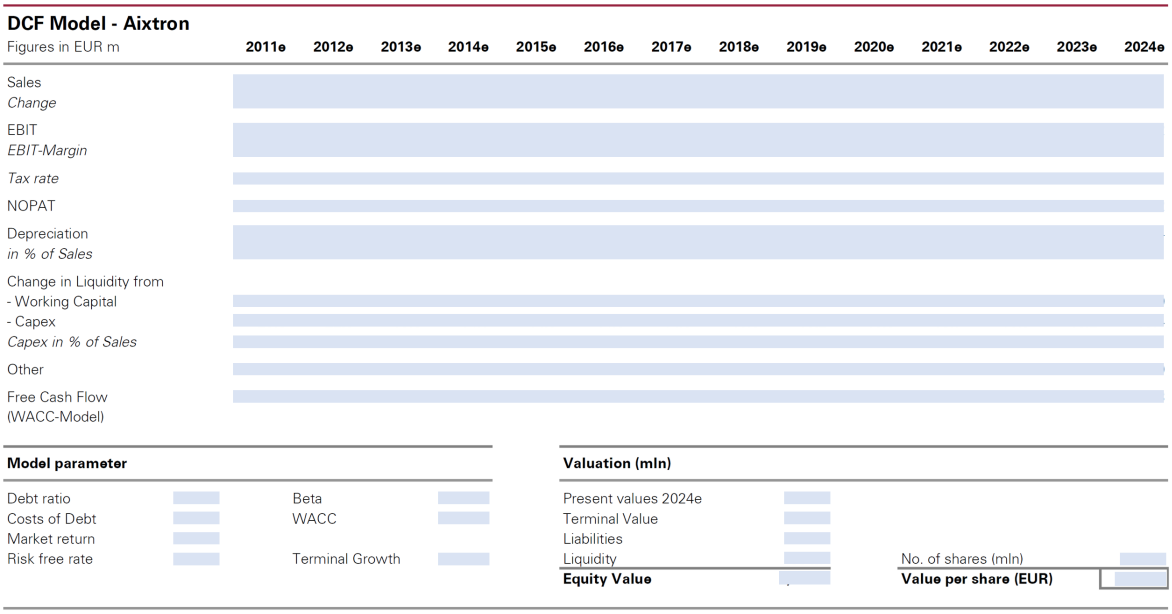


Figure B.2: Example of complete DCF equity report This figure shows a representative example of a discounted cash flow (DCF) valuation model table. This figure is taken from an equity report for Aixtron (Ticker = AIXGn), published by Warburg Research on October 27, 2011. For copyright reasons, all information in the table is redacted.

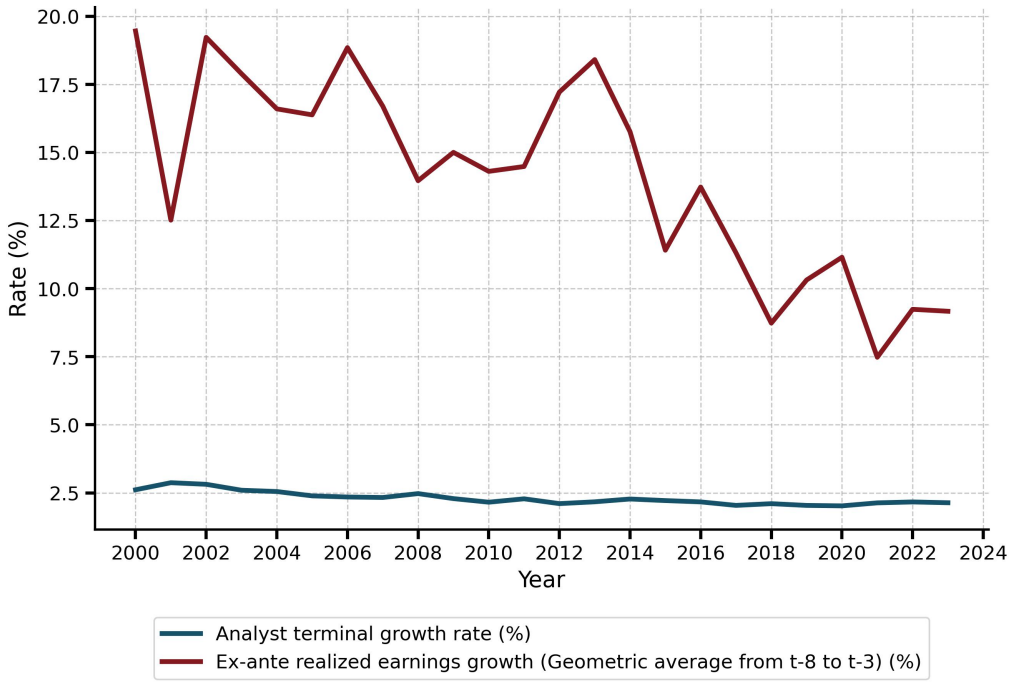


Figure B.3: TGR expectations versus historical returns This figure compares the time trends of historically realized earnings growth at the $t-8$ to $t-3$ horizon with analysts' terminal growth rate expectations at time t. Realized earnings growth corresponds to the geometric average realized between $t-8$ and $t-3$, averaged by year (t). Terminal growth rate is the average terminal growth rate expectation by year. The x-axis denotes years. The y-axis corresponds to rates, in percent (%).

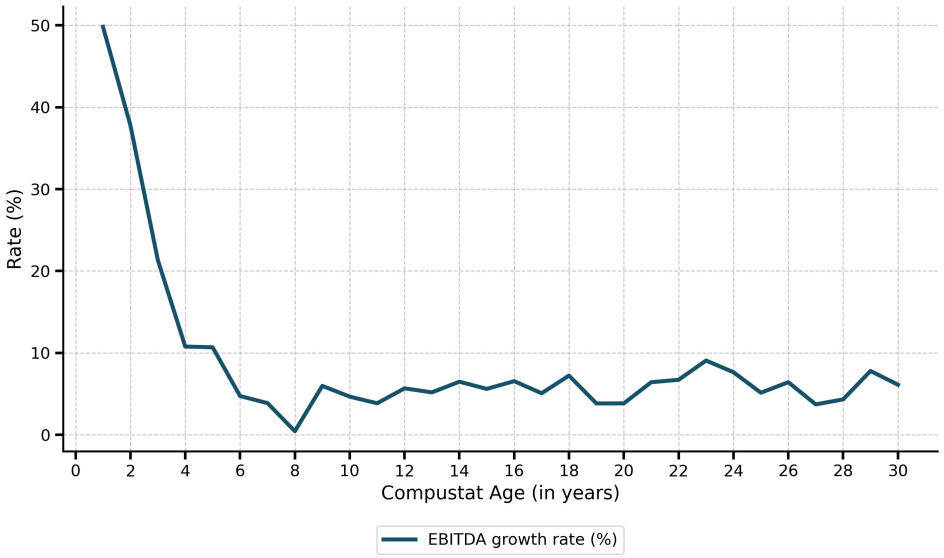


Figure B.4: Annual earnings growth rate and firm age This figure plots realized annual earnings growth as a function of firm age. The x-axis is firm age (based on the IPO date in Compustat). The y-axis is the average annual earnings growth rate across firms. The sample includes all firms included in Compustat for which the data item OIBDP (Operating earnings before depreciation) is available.

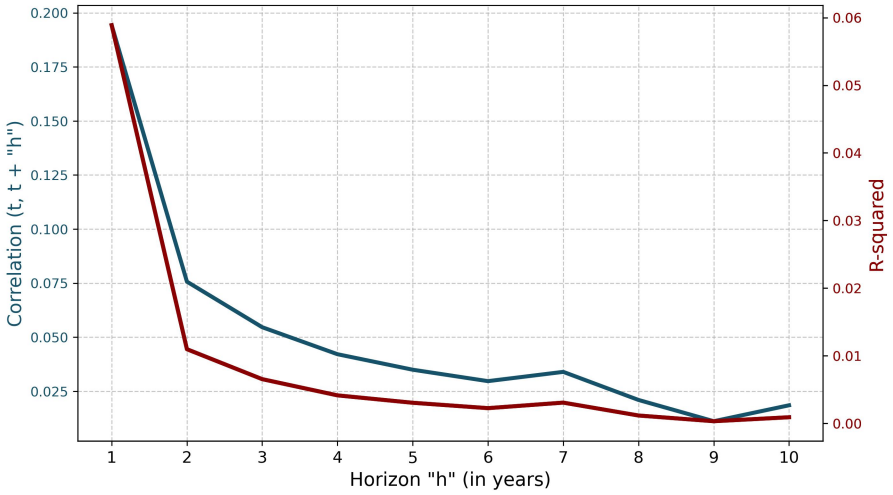


Figure B.5: Correlation between realized growth rates at different horizons This figure plots the average correlation between the realized earnings growth rate at time t and realized earnings growth rate at time $t+h$. Note both t and $t+h$ refer to realized growth rates observed after $t+h$. The blue lines shows the estimated correlation coefficients for different values of h. The red line shows the R^2 when regressing earnings growth at time $t+h$ on earnings growth at time t for different values of h.

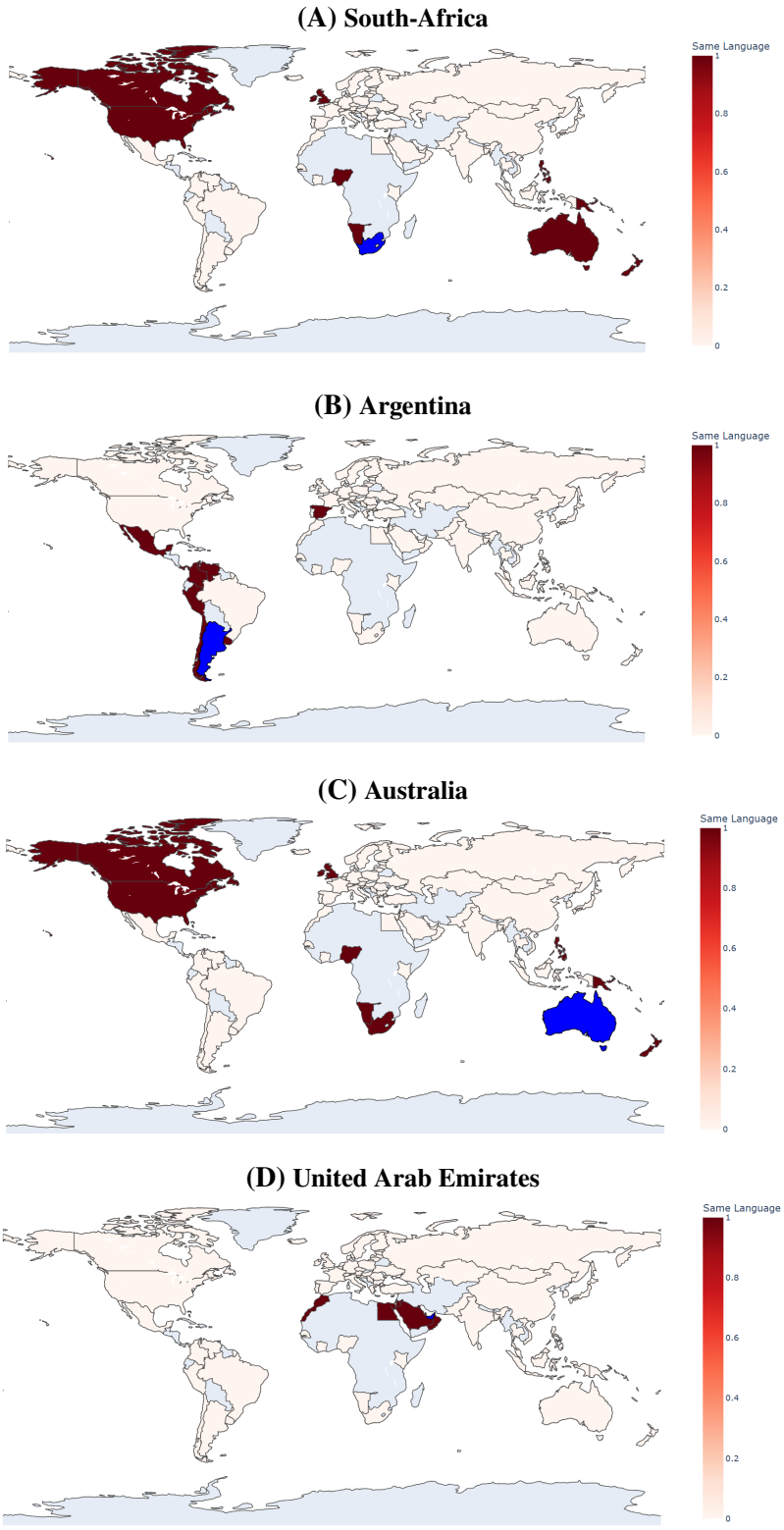


Figure B.6: Language similarity between countries This figure depicts language similarity between a focal country and other countries for four focal countries: (i) South Africa, (ii) Argentina, (iii) Australia, and (iv) the United Arab Emirates. We define a country's language based on its official language. For countries with multiple official languages, 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 prior to March 1, 2025), we use the most widely spoken language. The language similarity measure is assigned a value of 1 if the focal country and a given other country share the same defined language and 0 otherwise.

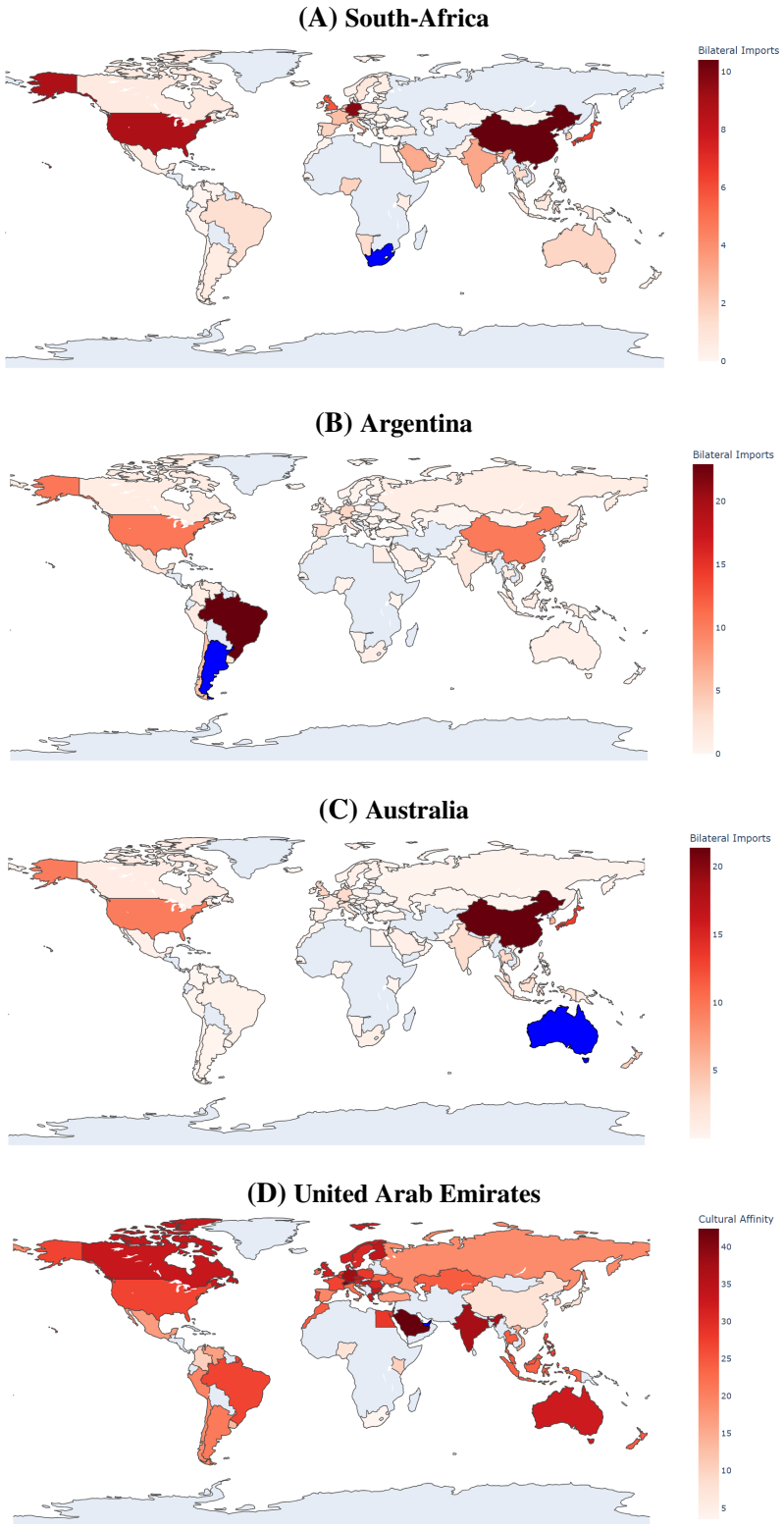


Figure B.7: Economies ties between countries This figure depicts economic ties between a focal country (in blue) and other countries, represented as the share of total bilateral imports. Four focal countries are considered: (i) South Africa, (ii) Argentina, (iii) Australia, and (iv) the United Arab Emirates. 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 the other country. 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 trade partners. The data, sourced from the OECD, cover the period 2000–2021. For each country pair, we compute the average share over this entire period. A value of 0 indicates no trading relationship on average, while a value of 1 signifies that the focal country traded exclusively with that single trade partner.

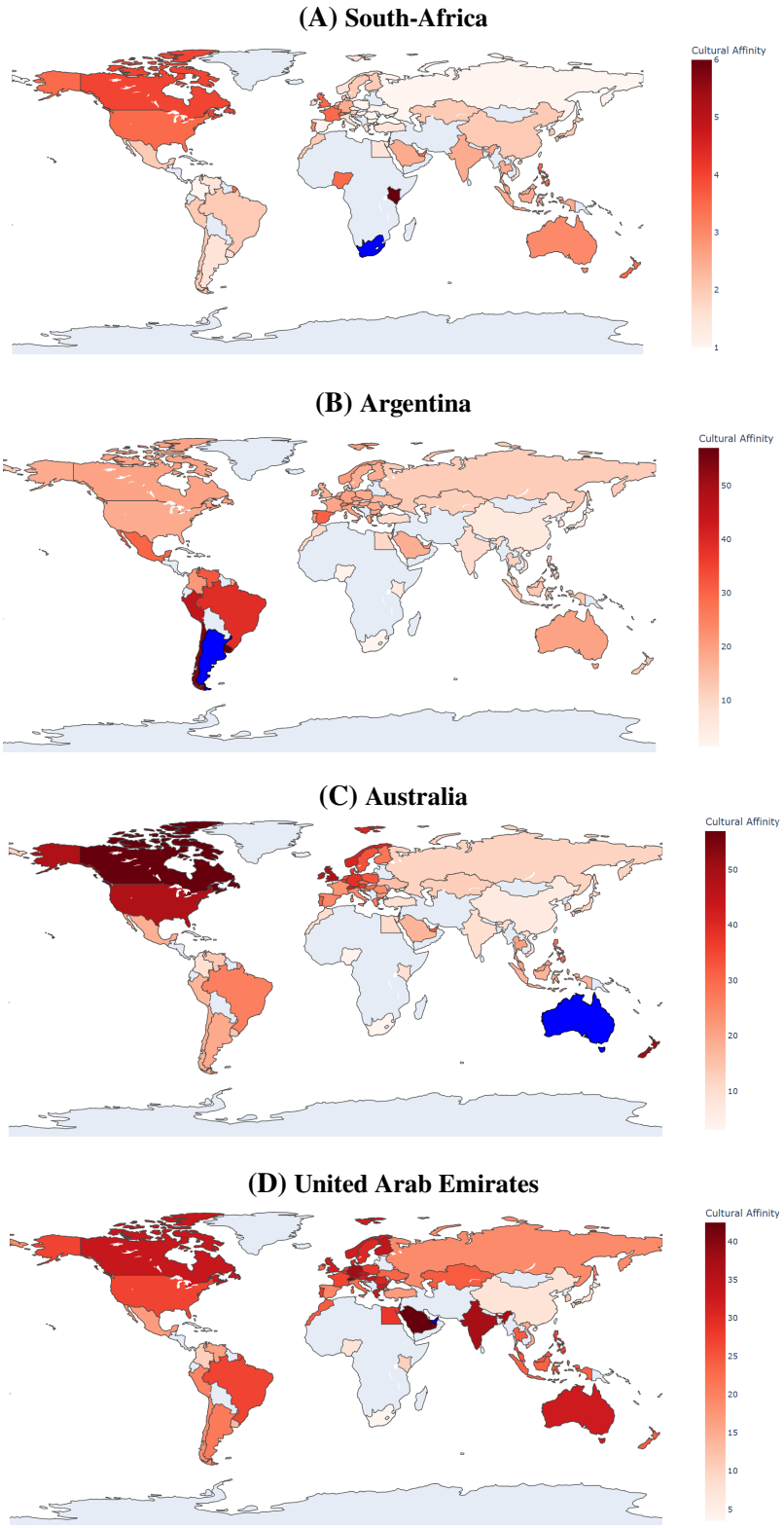


Figure B.8: Cultural affinity between countries This figure depicts cultural affinity between a focal country (in blue) and other countries, measured by the degree of overlap in music consumption. Four focal countries are considered: (i) South Africa, (ii) Argentina, (iii) Australia, and (iv) the United Arab Emirates. Music consumption overlap is measured 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.

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 level. The sample period is 2000–2023.

	Cosine similarity$_{a,b,i,t}$		
	Different First NAICS Digit	Same First 3 NAICS Digits	Same Firm
	(1)	(2)	(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 key words representing LDA topics. Words are selected based on reversed frequency weights, where a word's weight in a given topic is scaled by its relative frequency in the texts, similar to the strategy in Bybee et al. (2023b).

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	Consensus	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 discussions (Alternative topic model with 150 topics) This table presents the relationship between differences in textual discussion and disagreement in TGR expectations between pairs of analysts, using an alternative LDA model with 150 topics. The unit of observation is at the analyst a, analyst b, firm i, and forecast year t level. The sample period is 2000–2023. The dependent variable, $Disagreement_{a,b,i,t}$, is defined as the squared difference in TGR estimates between analysts a b, scaled by the absolute value of their sum: $(TGR_{a,i,t} - TGR_{b,i,t})^2/(|TGR_{a,i,t} + TGR_{b,i,t}|\ /2)$. The key independent variables of interest are the Jaccard and Cosine similarity scores, which capture the similarity of the analysts' textual discussion 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: Aggregate informativeness of TGR expectations This table replicates the results presented in Table 2, but uses aggregated measures of TGR expectations and earnings growth (i.e., aggregating across firms in each year). The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent. 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 expectations This table presents the relationship between analysts' trust toward other cultures and their TGR expectations about firm located in these countries. The unit of observation is at the firm i, analyst j, and forecast year t level. The sample period is 2000–2023. The independent variable, $Trust_{c,k}$, is equal to the value of the measure of "trust" averaged across all the participants in the data from the Eurobarometer survey "ZA Study Number" 2935, which was collected in 1997. The sample used in the regressions covers all firm HQ-country analyst-country pairs 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: Variance decomposition of TGR expectations (OLS) This table presents the explained variation in OLS regressions of TGR expectations on brokerage house, analyst, and firm fixed effects. Panel A replicates the specification in Column (1) of Panel A from Table 4 (i.e., full sample analysis, excluding firms, analysts, and brokerage houses with fewer than 2 observations in total, as their respective fixed effects fully account for the variation in TGR expectations). Panel B excludes brokerage houses with fewer than 2 observations, firms with fewer than 10 observations, and analysts with fewer than 50 observations, replicating the specification in Column (4) of Panel B in Table 4. Both panels also exclude brokerage houses that are not associated with equity analysts who switched employers during the sample period (2000–2023), as it is not possible to separate the effects of these brokerage houses from those of 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.7: Decomposition of analyst fixed effects (OLS) This table presents the explained variation in OLS regressions of the analyst fixed effects estimated in Table 7 on analyst characteristics. Panel A excludes firms, analysts, and brokerage houses with fewer than 2 observations in total, as their respective fixed effects fully explain the associated variation in terminal growth rates. Panel B presents robustness excluding cases with fewer than five observations. Both panels also exclude brokerage houses that are not associated with equity analysts who switched employers during the sample period (2000–2023), as it is not possible to separate the effects of these brokerage houses from those of their analysts (Abowd et al., 1999). The unit of observation is at firm i, analyst j, and forecast year t level. 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 a master's degree, MBA, or other degrees. Birth cohort is a series of birth year indicators. Analyst country is a categorical variable indicating an analyst's country at the time of producing the report. Variable definitions appear in Appendix A.

Panel A: Column 1 of Panel B in Table 7	Analyst FE$_j$						
	(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
Panel B: Column 6 of Panel B in Table 7	Analyst FE$_j$						
	(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