What Drives Very Long-Run Cash Flow Expectations?*

Paul H. Décaire† and Marius Guenzel‡

September 12, 2024

Abstract

We study the very long-run expectations of professional forecasters using a large novel dataset that combines forecasters' beliefs about firms' *terminal growth rate (TGR)*, qualitative text discussions of their beliefs, and detailed forecaster demographic information. We establish four findings about forecasters' TGR beliefs—their expectations about firms' long-run growth rate "until infinity." First, TGR expectations contain distinct economic information compared to shorter forecasting series and robustly predict realized long-run firm growth. Second, consistent with firm life-cycle models, a rich set of factors explains TGR beliefs for young firms, while country and industry factors account for most of the variation for mature firms. Third, there exists large, persistent heterogeneity in TGR expectations across forecasters, with forecaster demographics explaining a sizable share of this heterogeneity. Lastly, we use forecasters' textual discussions to estimate a topic model and open the black box behind quantitative expectations. Divergent long-term growth expectations stem primarily from subjective interpretations of similar topics, rather than from coverage of different topics.

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

Keywords: Long-term expectations, behavioral finance, valuation, professional forecasters, analyst identities and backgrounds, firm life cycle

*For helpful comments and suggestions, we thank Francesco D'Acunto (discussant), Spencer Kwon, as well as participants at the 2024 Mid-Atlantic Research Conference in Finance, the 2024 Behavioral Finance Conference at Bocconi, and the 2024 CESifo Venice Summer Institute Expectation Formation. For their excellent research assistantship, we thank our research team: Eisa Alloughani, Sebastian Gortarez Ortiz, Tej Kulkarni, Aditeya Pattiwar, Andre Radic, Reed Simons, Cole Streitmatter, and Mingyang 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

Long-run expectations are an essential ingredient in firm valuation and asset pricing. For example, the average public firm is expected to produce 70% and more of its discounted cash flows beyond a five- to ten-year horizon (Mukhlynina and Nyborg, 2018; Decaire and Graham, 2024). Moreover, even small variations in long-run beliefs imply large changes in valuations.[1] Yet, evidence is lacking on how market participants form expectations over long horizons, what factors influence their long-term beliefs, and the extent of similarity or dissimilarity in long-term beliefs across individuals.[2] Simply put, "collecting more data on long-run expectations would be useful" (Nagel, 2021), and would help inform and discipline models of belief formation and dynamics.

This paper provides direct evidence on the very long-term expectations of professional market participants. We employ a novel dataset with (1) more than 41 thousand forecasts by professional analysts of the *terminal growth rate (TGR)*—the growth rate they expect firms to sustain in the long run "until infinity"—for a large set U.S. and international firms, (2) analysts' qualitative text discussions of their beliefs, and (3) detailed analyst demographic and other background information. We document, among other things, that long-term TGR expectations only weakly correlate with shorter-horizon forecasting series and contain distinct economic information, that there exists large and persistent heterogeneity in TGR expectations across forecasters, and that both varying topic foci in analysts' written discussions and background heterogeneity strongly predict the TGR belief variation. More broadly and speculatively, our findings suggest that accounting for individual heterogeneity is particularly important for distant, and thus more abstract and malleable, economic contexts.

Our analysis utilizes two key building blocks. The first is the dataset on analysts' long-term TGR expectations, containing 41,429 terminal growth rate estimates, each accompanied by a qualitative discussion in the equity report by the forecasting analyst about the firm's prospects, averaging 800 words. The dataset covers 8,000 firms located in 49 countries

[1] For a simple example, consider a \$1 perpetual cash flow growing at rate g, to be discounted at 10%. A small, 1 percentage point change in g from, e.g., 1% to 2%, will result in a comparatively large, 12.5% change in valuation.

[2] Data limitations have constrained researchers to mainly focus on firms' short- and medium-term expectations, in particular those for one- to five-year horizons. We review this literature below.

between 2000 and 2023. To assemble it, we extract information from the original analyst equity reports using a series of pdf extraction functions, optical character recognition (OCR), and artificial intelligence document processing programs.

The second key building block of our analysis is the collection of comprehensive background information on the forecasting analysts. In our TGR dataset, we directly observe various pieces of identifiable analyst information, including the full name of each forecaster and their employer. Using this information, we merge our dataset with the universe of individuals included on LinkedIn.[3] This allows us to obtain detailed data on forecaster backgrounds, such as employment history, educational background, forecaster age, location of formative years, place of residence during formative years (high school and college), gender, and race. In total, we collect detailed personal information for 1,667 individuals located in 44 countries.[4] Moreover, a subset of forecasters in our data switch employers, allowing us to tease apart individual and firm effects in the analysis (Abowd et al., 1999).

We organize our analysis into two main parts. In the first part, we establish the systematic differences of long-term TGR expectations from shorter-term expectation series, along with their significant real economic information content.

In particular, TGR expectations are systematically different from the focal variable of prior literature, the IBES LTG measure of analysts' 3-5-year earnings growth forecasts. The average difference in TGR and LTG expectations is 10 percentage points (pp), as first shown in Decaire and Graham (2024). Moreover, as we document, the correlation between both expectation series is noticeably weak ($\rho = 0.08$). The considerable disconnect between both series indicates important differences in their information content, and a limited capacity of the 3-5-year LTG forecasts in capturing truly long-run beliefs of financial market participants.

Additionally, TGR expectations robustly forecast firms' realized long-run growth. We estimate an economically and statistically significant positive relation between analysts' TGR forecasts and actually realized firm growth over various longer-term horizons, with the

[3] The LinkedIn data, containing more than 750 million unique individuals, is obtained from Revelio Labs.

[4] By contrast, existing datasets on firm growth forecasts such as IBES lack any personally identifiable information about forecasters, making it impossible to link the variation in forecasts across analysts to underlying forecaster characteristics and experiences. For example, Bouchaud et al. (2019) are able to investigate variation in forecast stickiness across analysts as a function of the number of years an analyst is included on IBES or has provided forecasts for a specific firm or industry, but they are unable to link the variation to any personal characteristics of analysts.

estimated regression slope ranging between 0.5 and 1 depending on the examined horizon of realized firm growth. In the cross-sectional average over time, average TGR expectations predict a significant share of future realized average firm growth (R^2 ranging between 0.34 and 0.60). Since the correlation between analysts' TGR forecasts and the subjective discount rates they apply in the *same* reports is close to zero (Decaire and Graham, 2024), this implies that long-term growth expectations contain independent information that is useful in forecasting realized firm growth. Consistent with this, we also find that TGR expectations respond to increases in firms' default risk.[5]

In the second part of the analysis, we investigate three key dynamics: (i) how firm characteristics shape long-term TGR expectations throughout a firm's life cycle, (ii) the role of individual characteristics, backgrounds, and experiences in explaining heterogeneity in long-term belief formation dynamics across analysts, and (iii) how differences in the qualitative topic discussion in the equity reports, i.e., differential attention allocation, relate to TGR dispersion across analysts.

To guide the analysis, we first perform a variance decomposition 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 forecasting analyst), and brokerage house FEs (i.e., FEs for each analyst employer). Jointly, these fixed effects explain more than 65% of the TGR variation. Perhaps surprisingly, brokerage house fixed effects only negligibly contribute to the explained variation (approximately 2%), implying that organizational norms and protocols play a limited role in determining analyst expectations. Both firm fixed effects and analyst fixed effects contribute substantially to the explained variation. The significance of individual analyst fixed effects, after partialling out brokerage house effects, implies persistent heterogeneity in TGR expectations across forecasters,[6] and motivates our closer focus on heterogeneity in analyst backgrounds below.

We next investigate further the influence of firm characteristics. We find that the

[5] We also find that forecasters' long-term TGR stock growth expectations are systematically below historically realized long-term growth rates in every year of our sample. We discuss possible underlying reasons in Section 4.3. These patterns add to the differences between the TGR expectation series in our analysis and LTG expectations studied in prior work, as the latter are systematically above historical growth rates (Nagel and Xu, 2022; Bordalo et al., 2023).

[6] This finding is consistent with prior work documenting persistent individual heterogeneity in macro-finance beliefs of retail investors (Giglio et al., 2021).

explanatory power of firm fixed effects varies significantly over the firm's life cycle, being more important in younger firms and gradually declining with firm age. As firms mature, firm country and industry factors increasingly account for a larger share of TGR expectations, reaching 60% and more for firms aged 70 years and older.[7] Furthermore, the within-firm, within-year dispersion in analyst TGR expectations declines close to monotonically with firm age, consistent with a firm life cycle component influencing analysts' expectation formation.

We then examine further forecaster characteristics and the large, persistent heterogeneity in TGR expectations across analysts. We formally estimate analyst fixed effects using a mover design (Abowd et al., 1999), exploiting analysts who switch brokerage houses in the sample, and then regress the estimated fixed effects on forecaster observables. Forecaster demographics can explain as much as 20-30% of the variation in analyst fixed effects, which is substantial in comparison to prior work. For example, in Giglio et al. (2021)'s surveys of wealthy U.S. retail investors, respondent characteristics can only explain 3-10% of the estimated individual fixed effects.[8] One likely reason for our increased ability to explain forecaster expectation heterogeneity is the larger dissimilarity in forecaster characteristics and backgrounds in our sample, e.g., encompassing analysts from many different countries. Significant factors that contribute to the explained variation in our setting include the age (birth year) of forecasters, their race, and, indeed, their geographic location (country).

We further document, beyond the presence of persistent heterogeneity in TGR forecasts, strong evidence of local extrapolation in very long-run beliefs, consistent with recent work in other domains (e.g., Kuchler and Zafar, 2019). Recent GDP growth in an analyst's country of residence strongly predicts their TGR expectations for firms located in *other* countries, controlling for GDP growth of the firm's country. This pattern is unlikely to be explained by common shocks to a cluster of neighboring countries affecting expectations, or special

[7] This finding relates to Decaire and Graham (2024) showing that analysts systematically consider macroeconomic variables—GDP growth—when determining TGR forecasts, with macro-variable series explaining 92% of TGR expectations when averaged at the country-year level.

[8] Other prior work that uncovers robust statistical associations between demographics and expectations commonly finds relatively low explanatory power of demographics for the variation in expectations as well (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). These papers also examine samples that compared to ours are more homogeneous along important dimensions, such as focusing on households from a single country (e.g., the U.S. or one of the Scandinavian countries).

linkages between the analyst country and the firm country, as the granular nature of our data allows us to include firm-country*analyst-country fixed effects. Instead, consistent with cross-country extrapolation, the patterns are fully driven by observations where analyst countries are located at a considerable geographical distance from firm countries.

In a final step, we link analysts' quantitative TGR expectations, and heterogeneity therein, to differences in the thematic focus and tilt in the text portions of the equity reports. As such, ours is a rare setting where we observe both quantitative expectations and qualitative text where analysts describe and discuss their quantitative estimates, allowing us to open the "black box" behind the belief formation process.

To this end, we estimate a latent Dirichlet allocation (LDA) model, and investigate how professionals allocate topic focus on both the extensive and intensive margin. At the extensive margin, the top five topics in a report account, on average, for 84% of the entire text content. At the same time, for the average analyst pair, only 43% of all topics covered by at least one analyst of the pair are included in both reports. Together, these patterns indicate the existence of person-specific factors partially shaping the decision to focus attention on a small subset of topics (cf. Gabaix, 2014). At the intensive margin, we find material differences in the weights analysts allocate to each covered topic. In our pairwise sample, we estimate a cosine similarity of 0.8, i.e., a *dissimilarity* of 0.2, for the extent to which the analyst pair apply different weights to their overlapping topics.[9]

We then explore how the variation in analysts' textual discussions predicts the differences in their long-term expectations. Exploiting the fact that we frequently have multiple analysts evaluating the same firm at the same time, we continue to employ measures of pairwise topic focus differences: Jaccard[10] and Cosine dissimilarity. We find that prioritizing overlapping topics differently, rather than discussing different topics, is the main predictor of heterogeneity in the quantitative TGR expectations.[11] Working with the pairwise extensive- and intensive-

[9] Cosine similarity can take values between -1 and 1. A value of 1 indicates that both analysts put exactly the same weight on each overlapping topic, whereas a value of 0 would suggest no relation. Across reports, the average topic consumes 15% of the overall qualitative discussion.

[10] Jaccard similarity quantifies, at the extensive margin, the overlap between the sets of topics included in both reports, with its value ranging from 0 to 1.

[11] In ongoing work in progress, Bastaniello, Decaire, and Guenzel (2024) show, among other things, that different aspects of topic attention allocation matter for long-term versus short-term valuation-relevant variables, indicating a discrepancy in the qualitative factors that are influential at different horizons.

margin measures allows us to collapse the high dimensional text data space into a single easy-to-compare number for each analyst pair, and offers a new angle to explore the role of qualitative information in shaping expectations.

Lastly, we connect the qualitative topic focus results with our extrapolation evidence. Greater geographic distance between analysts is positively associated with greater disagreement on the intensive margin. Additionally, TGR belief variation are particularly shaped by differences in the topic intensity with respect to geographic and macroeconomic topics. These findings indicate that local conditions appear to influence professionals' coverage intensity, serving as one underlying mechanism for extrapolation.

Overall, we present the first comprehensive analysis of the origins and drivers of professional forecasters' very long-run cash flow expectations. The findings we establish can help guide macro-finance models. One key feature is large, individual heterogeneity in long-run beliefs, likely amplified in long-horizon forecasting contexts, where predictions cannot be rapidly compared with realized outcomes and are thus more malleable. A second important model ingredient is extrapolation, and as our heterogeneity and topic analysis results reveal, specifically extrapolation from attending differently to local information and signals.

Our paper makes several literature contributions. First, we contribute to the literature on the macro-finance belief formation process of investors and financial market participants (e.g., Giglio et al., 2021), and the literature on the effect of personal characteristics, experiences, and extrapolation on beliefs, learning and economic outlooks (e.g., Malmendier and Nagel, 2011, 2016; Kuchler and Zafar, 2019; Hartzmark et al., 2021; cf. also the literature referenced in footnote 8). By focusing on individual analysts' beliefs and backgrounds, we also add to the literature in behavioral corporate finance, in particular the less-developed strand focusing on "third parties," i.e., non-investors and non-managers (cf. Malmendier (2018) and Guenzel and Malmendier (2020) for recent surveys of the field of behavioral corporate finance).

Furthermore, we advance the literature that studies investor and analyst expectations beyond very short horizons, yet typically examines LTG expectations from IBES that capture medium-term (3-5-year) forecasting horizons (Nagel and Xu, 2022; De la O and Myers, 2023; Bordalo et al., 2023). We also contribute to an emerging literature that studies the valuation practices of professional forecasters using directly observable valuation inputs (Decaire,

2023; Decaire and Graham, 2024; Decaire et al., 2023). Finally, our results contribute to a burgeoning literature that combines quantitative with qualitative information (e.g., text data that underlie quantitative forecasts) to shed light on individuals' belief formation process (Kendall and Oprea, 2023; Demgensky and Fritsche, 2023; Bybee et al., 2023b; Andre et al., 2024; Charles and Kendall, 2024).

The remainder of the paper proceeds as follows. Section 2 provides institutional background. Section 3 introduces the methodology and data. Sections 4 and 5 present the results. Section 6 concludes.

2 Institutional Details

We assemble the long-run cash flow expectations dataset from equity reports published by professional analysts (i.e., from the original documents). Hence, it is useful to first provide 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 (Graham and Harvey, 2001; Mukhlynina and Nyborg, 2018) to perform valuation, with one or both strategies being included in close to 100% of equity reports housed by our data provider, Refinitiv. Decaire and Graham (2024) also shows that 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. Although the valuation strategies share similarities, our data collection focuses on reports produced with DCFs for two institutional reasons.

First, in contrast to valuation multiples that only disclose a single metric—the multiple—DCFs require analysts to report detailed modeling assumptions, allowing us to clearly distinguish the terminal growth rate used by analysts from the other parameters of the valuation model. Appendix Figure B.1 presents an example of a complete DCF model from an equity report.[12] This granularity of the data guarantees that we never have to estimate or

[12] For copyright reasons, we have redacted the numbers used in the model.

approximate any of the terminal growth rates. Instead, our analysis is conducted entirely with data directly observed in the analyst reports.

Second, many past studies using analyst data have relied on commercial databases, such as IBES or Value Line, which collect and provide mainly short- and medium-term expectations. As we also discuss in detail in Section 4.1, the "long-term" growth estimates reported in these databases (e.g., LTG) do not reflect truly long-term expectations provided by analysts. LTG from IBES and Value line reflect forecasts made for a three- to five-year horizon, while our data capture analysts' *very* long-term expectations. The data limitations present in existing datasets have forced researchers to rely on potentially inexact proxies when determining terminal growth rates in their analysis. For example, Figure 3, which we discuss in more detail in Section 4.1, shows that the annual LTG from IBES averages 13.03%, with the interquartile range taking values between 5.50% and 17.0%, an order of magnitude larger compared to the average terminal growth rate in the equity reports of 2.2%, and an interquartile range between 1.5% and 3.0%. Consequently, if researchers directly relied on the annualized LTGs from IBES as a proxy for the long run, they would systematically overstate long-term growth rate expectations.

We note that analysts disclose DCF modeling details in their equity reports on a voluntary basis. Studies have found that the extent of information disclosure of the DCF model assumptions is positively associated with report accuracy (Asquith et al., 2005; Hashim and Strong, 2018), and more detailed information disclosure leads to larger market reactions following changes in recommendations (Huang et al., 2023). By detailing their valuation thesis, informed analysts have the opportunity to differentiate their work from their uninformed rivals, gaining credibility in the process. In total, this suggests that our data differs advantageously from datasets that unconditionally collect earnings forecasts, as our data is likely to be sampled from a more informed subset of analysts.

3 Methodology and Data

3.1 TGR Forecasts from Analyst Equity Reports

We initially downloaded 157,549 equity reports with mentions of DCF from 55 major equity research firms. We restricted the time window to reports published in the first quarter of the calendar year (January 1^{st} to April 1^{st}) from 2000 to 2023, to ensure that our data items were systematically measured at a similar time point during the year. In cases where analysts published more than one report on the same firm during those months, we systematically kept the earliest publication of that calendar year to avoid duplicates for a given analyst-firm-year pair. This procedure resulted in 41,429 reports that (1) contained the terminal growth rate numerical estimated and, crucially for our goal of examining driver of the heterogeneity in long-run expectations across analysts, for which we (2) could unambiguously identify the *analyst name* and *location* at the time of the report, the latter by using their landline country indicators.

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

3.2 Firms and Coverage

The final sample of 41,429 equity reports, coming from 41 of the largest brokerage houses operating throughout the world, covers 8,010 firms located in 49 countries, and include analysts located in 44 countries during the 2000-2023 period (cf. the summary statistics in Table 1, which we discuss in more detail in Section 3.5). The average forecasted firm owns assets with a book value of $15.8 billion USD and has an investment rate of 5.55%. These magnitudes are comparable to the IBES universe, where the average firm has assets with a book value of $15.2 billion USD, and an investment rate of 5.3%. Overall, this suggests that the firms in our sample are of comparable size and invest with similar intensity as those included in other broadly distributed commercial analyst datasets.

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

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

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

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

3.3 Analyst Backgrounds from Universe of LinkedIn Data

We merge the TGR dataset on long-run expectations described in Section 3.1 with the universe of LinkedIn users obtained from Revelio Labs, thereby obtaining additional (self-reported) personal characteristics of forecasters, including their educational background and (approximate) age. The merge is made possible since as described, the TGR dataset contains

key identifiable information on the forecasting analysts, including their full name and the name of their employer, allowing for a clean and unambiguous merge with the Revelio data.

In linking the datasets, we require the names across individuals in the TGR and LinkedIn datasets to match precisely. To further err on the side of caution in constructing the matched dataset, we also require the LinkedIn-provided employment history to match the analyst's employer in the equity reports,[13] as well as the name-employer matching to result in a unique LinkedIn user that is matched to the data.[14] This procedure leads to 1,667 uniquely matched analysts across the TGR and LinkedIn datasets (Panel A of Table 1) out of the 5,300 equity analysts included in our sample (31%).

3.4 Latent Dirichlet Allocation Estimation

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

3.4.1 Text Pre-Processing

We start by extracting the text from each report using a series of Python routines. Specifically, in this step, we *exclude* information contained in figures, tables, and the legal boilerplate disclaimers at the end of reports. The objective is instead to capture the valuation thesis that professionals articulate to support their numerical analysis. In a final step, we do a final pass on the extracted text to address extraction issues that might have been missed by the Python routine by using Meta's LLM, Llama-2-70b.[15]

Then, we apply five transformations to the text data. First, for an equity report to be included in the LDA estimation, it must contain at least 100 words. Second, we remove verbs from the text. Third, we remove stopwords and lemmatize words using standard Python libraries. Fourth, we organize words into unigrams and bigrams, based on the terms'

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

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

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

occurences in our full set of reports. Pairs of words that often arise together receive higher likelihood of being treated as bigrams than as two unigrams.[16] Last, for a unigram or bigram to be included in our final dataset, the term has to appear at least once in 0.1% of all the reports used in the analysis.

3.4.2 LDA Model Estimation

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

To estimate an LDA model, one must select three hyper-parameters: (i) the number of topics (K), a parameter governing the distribution of topics within reports (α), and (iii) a parameter determining the distribution of terms within a topic (η). We set α to be equal to 0.05, and η to 0.02. These values were obtained when optimizing the perplexity score of our selected model. These magnitudes are also broadly consistent with the estimation standard for LDA models, where a simple rule of thumb such as $\frac{1}{\text{Number of topics}}$ is generally a reasonable starting point for the estimation.[17]

To determine the optimal number of topics (K), we estimate different versions of the LDA model using from 10 to 200 topics, in increments of 10 topics. Our model selection strategy for the number of topics is based on three metrics: (i) the coherence score, (ii) the perplexity score of the LDA model, and (iii) a manual investigation of the keywords contained in each topic. The coherence score provides a measure of the topic model's interpretability by comparing the semantic similarity of the words included in each topic. A higher coherence

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

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

score indicates greater semantic similarity across the words in each topic. The perplexity score indicates how well the model describes our set of documents, with a lower perplexity score suggesting a better fit of the data. While these metrics are useful in guiding model selection, we also manually examine the set of words included in each topic to confirm their *natural* interpretability.

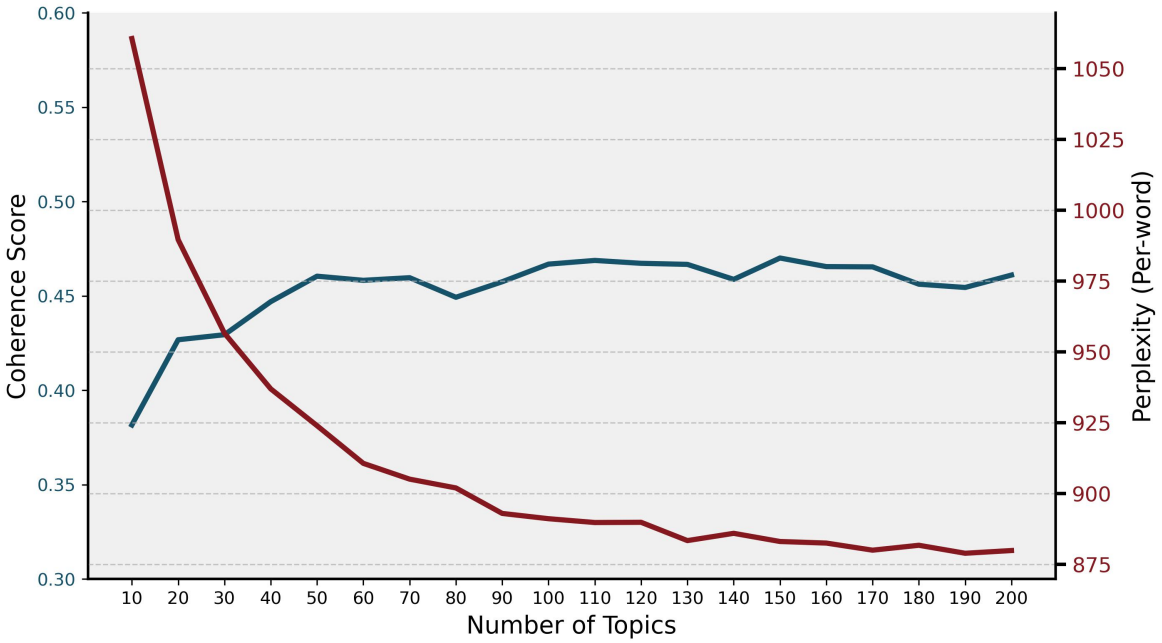


Figure 1: Coherence and Perplexity Scores This figure plots the coherence and the in-sample perplexity scores when using a different number of topics to estimate the LDA model.

Figure 1 plots the coherence and perplexity scores over the range of possible numbers of topics included in the LDA (Appendix Table B.2 reports the keywords included in our favored model.) As the figure shows, most of the improvements in coherence and the in-sample perplexity scores are obtained when reaching 50 topics, and after that improvements level off. Upon manual inspection, we also find that the 50-topic model is parsimonious, and contains topics that are meaningful, mostly non-overlapping, and distinctive.

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

3.5 Summary Statistics

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

In Panel B, the average TGR forecast in the data is 2.2%. Figure 2 plots the average TGR across firms and analysts over time, showing that average long-run expectations have somewhat declined over time, from around 2.8% at the beginning of the sample in 2000 to closer to 2.0% towards the end of the sample. In terms of levels, the TGR expectations are an order of magnitude smaller than the average IBES-based LTG expectation for the sample firms of 13.0%. (We will contrast TGR and LTG expectations further in Section 4.1.) Also shown in Panel B of Table 1, about 1 in 3 equity reports and TGR forecasts come from foreign analysts, i.e., analysts who forecast the long-run growth of a firm headquartered in a different country than where the analyst is located.

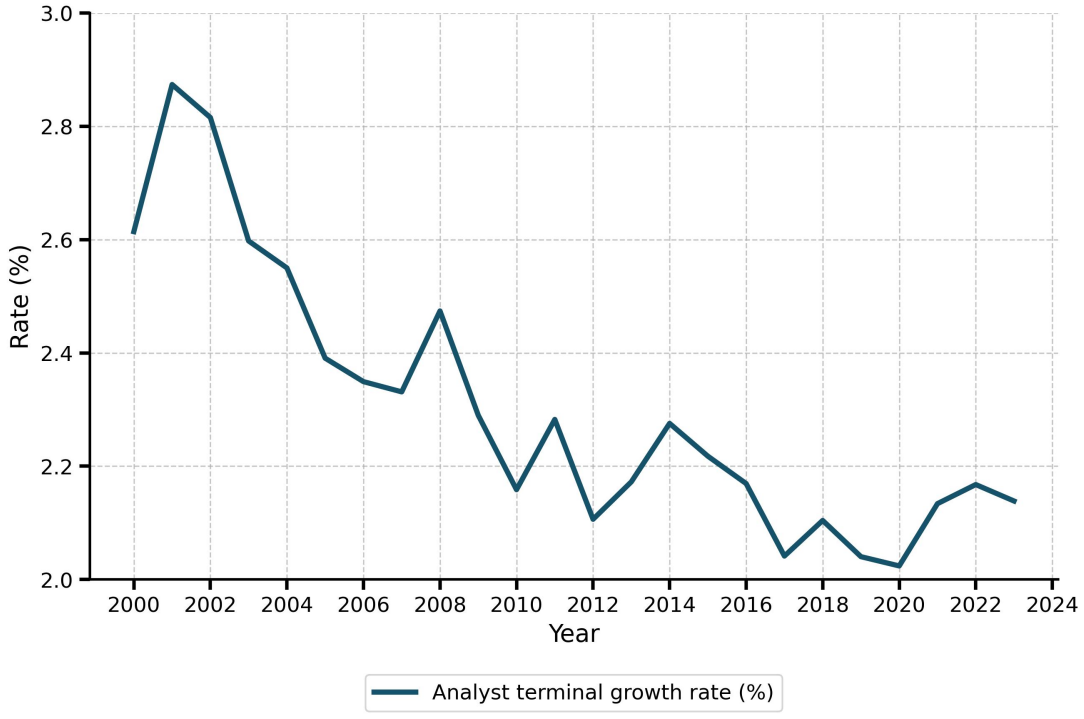


Figure 2: Average TGR expectations over time This figure plots the time trend of average terminal growth rates across analysts and firms over the sample period from 2000 to 2023.

Panel C contains demographic and other information about forecasters. Just under

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

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

4 Economic Information Contained in TGR Expectations

This section establishes the economic importance of analysts' long-term cash flow expectations, in three steps. First, we show that terminal growth rate expectations contain distinct economic information relative to other expectation series used in prior work. In particular, we show that terminal growth rates shares limited overlap with the existing long-term expectation benchmark used in the literature, IBES LTG—a measure that captures analysts' expectations over a 3-5-year horizon. Second, we show that terminal growth rate expectations predict ex-post realized long-term firm earnings growth, known in hindsight. Finally, terminal growth expectations respond to increases in firms' default risk.

4.1 TGR Versus IBES LTG Expectations

We start by examining the relation between terminal growth rates and the IBES LTG variable, and establish that both variables share limited similarities. Panel B of Table 1 reports that the correlation between the two series is weak ($\rho = 0.08$). Contrasting the time trends in

both variables, we find that the terminal growth rate yearly average has steadily declined over the sample period (Figure 2), whereas the average level of IBES LTG has remained constant. Further, comparing the levels of both series (Panel A of Figure 3), we observe a systematic and persistent gap between the two series, averaging 10.0 pp throughout the sample period, and showing that TGR expectations are substantially smaller than IBES LTG expectations. Lastly, Panel B of Figure 3 shows a scatter plot representing the relation between IBES LTG and terminal growth rates. Consistent with the weak correlation presented in Panel B of Table 1, it is difficult to visually discern a clear relation between both variables, given their considerable degree of dispersion. The figure also clearly shows that IBES LTG exhibits substantially more volatility than the terminal growth rate. Table 1 confirms the visual inspection of the scatter plot and shows that the standard deviation of IBES LTG (22.5 pp) is 17 times greater than that of the terminal growth rate (1.29 pp).

Adding to the differences between the TGR and LTG expectation series, in Appendix Figure B.2, we also find that forecasters' long-term TGR stock growth expectations are systematically below historically realized long-term growth rates in every year of our sample, whereas LTG expectations studied in prior work are systematically above historical growth rates (Nagel and Xu, 2022; Bordalo et al., 2023).

These differing patterns are likely explained, at least in part, with survivorship bias affecting realized growth series, especially over longer horizons, as also discussed in Decaire and Graham (2024). (We further examine how TGR expectations respond to firms' default risk in Section 4.3.)

Altogether, the evidence in this section establishes that TGR expectations are systematically different from other expectation series that target growth beyond a one- or two-year horizon, such as the IBES LTG series. This implies that our existing understanding of how professional forecasters and market participants form long-term expectations about firms' growth potential remains limited.

4.2 Predictive Power of TGR Expectations for Realized Growth Rates

To establish the economic information content of TGR expectations, as well as further validate this expectations measure, we study to what extent TGR beliefs predict firms' future realized

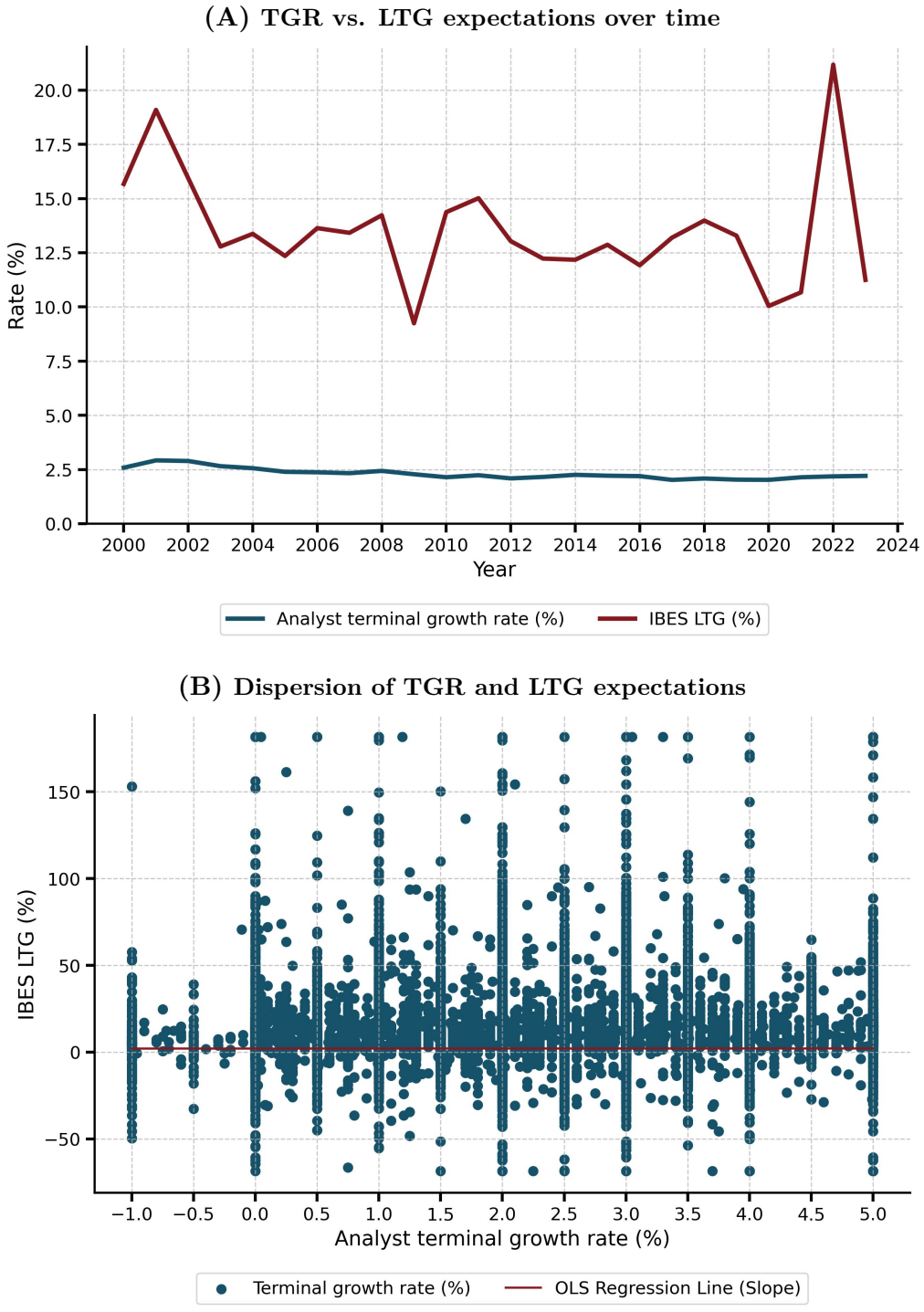


Figure 3: TGR versus IBES LTG expectations For Panel A, the x-axis is expressed in years, and the y-axis in percentage points. Panel A compares the time trend of IBES LTG expectations with analysts' terminal growth rates over the sample period. Panel B's x-axis represents analysts' terminal growth rates, and the y-axis corresponds to the firm-year LTG consensus (median) provided by IBES. Each blue dot denotes an observation. The red line is the slope from an OLS regression between IBES LTG and terminal growth rates. For illustration purposes and readability, we winsorize the LTG variable at the 0.1 and 99.9^{th} percentiles.

long-term growth.

We first reiterate that terminal growth rates are inputs from analysts' DCF models. Decaire and Graham (2024) show that the modal DCF model sets the explicit forecast horizon to be equal to three years and then applies a constant terminal growth rate to any subsequent cash flows. We note that this horizon has also been used in other contemporaneous studies (e.g., Hommel et al., 2023) that encounters methodological aspects similar to ours.

Against this background, in our main specification, our goal is to capture growth expectations from period $t+4$ and onward, when the modal analyst starts applying the terminal growth rate in the DCF model. Our main measure of realized growth is then the geometric average of the ex-post realized earnings growth from year $t+4$ to $t+8$, measured for each firm in our sample. Comparing the TGR beliefs with this ex-post measure of realized growth in Figure 4, we find that both series strongly comove together, but note differences in their respective levels (i.e., TGR expectations − ex-post realized growth = −3.2 pp). The fact that both series share a similar trend is reassuring, but the difference in levels calls for additional tests that we conduct at the end of this section.

To formalize the finding of a significant comovement in TGR expectations and realized firm growth rates from Figure 4, we run a linear regression where we consider three possible alternative ways to measure the ex-post realized growth: the average geometric earnings growth for each firm in our sample over the periods $t+1$ to $t+5$, $t+4$ to $t+8$, and $t+4$ to $t+13$. The last window expands the horizon over which we measure realized growth to account for a 10-year period, at the expense of a reduced sample size.

Panel A of Table 2 presents the results. Focusing on our preferred specification, $t+4$ to $t+8$ in Column (2), we find that a one percentage point increase in TGR expectations is associated with a 0.70 percentage points increase in the ex-post realized growth. The strong, positive relation between TGR beliefs and realized growth persists across the other examined realized growth horizons in Columns (1) and (3), respectively. Further, focusing on the annual average growth measures across firms in Panel B, we find that annual average terminal growth rates explain a significant share of ex-post realized growth ($R^2 = 0.44$).

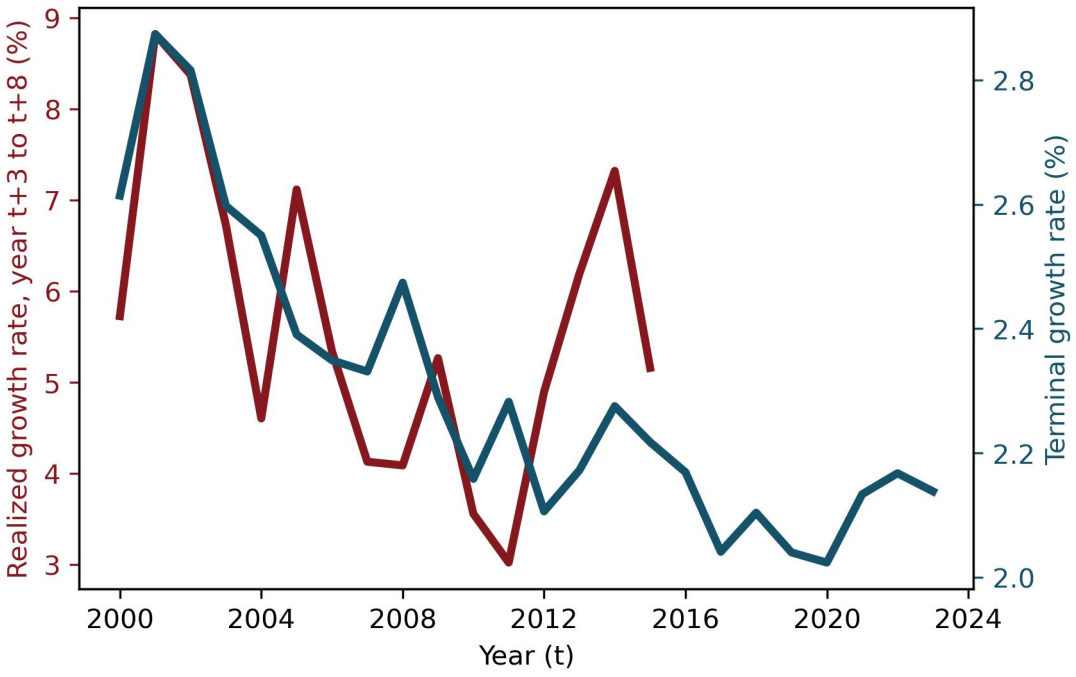


Figure 4: TGR expectations versus realized returns This figure compares the time trends of ex-post realized earnings growth at the $t+3$ to $t+8$ horizon with analyst terminal growth rates. Realized earnings growth corresponds to the geometric average realized from $t+3$ to $t+8$, averaged by year (t) in the sample. Terminal growth rate is the average terminal growth rate by year. The x-axis denotes years. The left-hand-side y-axis is associated with the ex-post realized returns, and the right-hand-side y-axis relates to the terminal growth rate.

4.3 TGR Expectations and Default Risk

We next investigate plausible factors that can help explain the gap between ex-post realized growth and the TGR measure of long-term expectations described in the previous section, focusing on the firm's probability of default. Naturally, ex-post realized returns can only be measured for surviving firms, i.e., those that did not go bankrupt over the sample period. However, ex-ante, analysts are likely to account for the probability of default, especially when forming expectations over very long horizons. For example, Decaire and Graham (2024) document that going back to 1950, close to 1 in 4 firms included in Compustat North America went bankrupt or were liquidated. For those firms, bankruptcy occurred after 9.4 years of operation on average.[18] These large liquidation probabilities highlight the need to account

[18] To obtain this statistic, they restrict their sample to firms that experienced bankruptcy, liquidation, or that were still active in Compustat as of August 28, 2023. For example, M&A targets, firms involved in going-private transactions, among others, were excluded from the calculation.

for default risk when forming long-term expectations.

To formalize this intuition, we construct an ex-ante measure of firms' probability of default following the naive Merton distance-to-default model discussed in Bharath and Shumway (2008) at two horizons, 10 and 20 years. We then examine the association with TGR expectations.

Table 3 presents the results. We include firm, country, and industry fixed effects, allowing us to evaluate how different levels of default risk impact the terminal growth rate for firms operating in the same country and industry during the same year. We also control for firm age, measured as the natural logarithm of the firm's post-IPO age, since our analysis below shows that firm age is one of the main drivers of TGR expectations. Using this specification, we estimate a negative link between TGR expectations and default risk, indicating that analysts appear to apply a 0.21 pp discount, a 10% reduction compared to the sample average, to firms that have a probability of default above the sample median. Considering that, conditional on being in the *high* risk of default group (above the median), the average probability of default is equal to 5.8%, the 10% reduction in the terminal growth rate is commensurate.

Altogether, the negative association between default risk and analyst TGR expectations, even when holding firm age constant, helps explain the higher level of realized firm growth in Figure 4, and further indicates the information content of TGR expectations, as analysts actively adjust their long-run forecasts in response to default risk variation.

5 Explaining Heterogeneity in TGR Expectations Across Firms and Analysts

Having established (1) that TGR expectations have distinct properties compared to other long-term expectations variables studied in the existing literature, and (2) that they predict realized long-run firm growth, we focus our efforts on understanding the key drivers of analysts' long-run TGR expectations. Specifically, we employ a variance decomposition of TGR beliefs, and link the (heterogeneity in) beliefs to a host of underlying firm as well as analyst characteristics.

5.1 Variance Decomposition

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

We consistently find that the three sets of fixed effects explain a large portion of between 60% and slightly above 70% of the variation in TGR expectations. Brokerage house fixed effects contribute only a very small share to the explained variation, whereas analyst fixed effects and firm fixed effects drive most of the explained variation. For example, in the largest sample in Column (1) of Panel A, brokerage house fixed effects contribute 2% of the explained variation, whereas analyst and firm fixed effects contribute 71% and 27%, respectively. Similarly, in the most narrow sample in Column (4) of Panel B, we estimate contributions of 15%, 52%, and 33%, respectively.

One limitation of the approach in Table 4 is that the ANOVA-based estimation is not indifferent to variations in the sequential inclusion order of the brokerage, analyst, and firm fixed effects. As an alternative approach, we therefore estimate standard OLS regressions in which we include the various fixed effects one at a time, without any other control variables. In each case, we will estimate an *upper bound* for the amount of variation that can be explained by the inclusion of the given fixed effects. Intuitively, the included fixed effects may be correlated with other (omitted) variables that also affect TGR expectations (cf. the related discussion in Ayyagari et al., 2008).

Using the alternative OLS approach, brokerage fixed effects alone can explain a very small share of only up to 1.4% of the total variation in TGR expectations. Analyst fixed effects alone can explain a much larger share of up to 49% of the TGR variation. Firm fixed effects alone can explain a similarly large share of up to 46% of the variation. It is also worth noting that brokerage and firm fixed effects jointly can explain up to 47% of the variation, whereas including in addition analyst fixed effects leads to an incremental R^2 of 0.20 (i.e., a total

R^2 of 67%).[19] These numbers imply that 20% is a *lower bound* for the variation in TGR expectations that can be explained directly by analyst fixed effects and indirectly by other variables (other than brokerage houses and firms) that analyst fixed effects predict.

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

5.2 Firm Characteristics

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

5.2.1 TGR Expectations Over Firms' Life Cycle

We first relate TGR expectations to firm age in Figure 5, which shows that forecasters' TGR expectations monotonically decline with firm age. The average TGR in the sample is 2.5% for young firms aged 10 years and less, and declines to 0.5% for very old firms, reflecting diminishing growth opportunities with firm age. These magnitudes are obtained after controlling for firm and year fixed effects. Columns (3) to (5) of Table 3 confirm that the decline in TGR expectations with increasing firm age is statistically significant.

Motivated by this evidence, we then ask when the explanatory power and importance of firm fixed effects are most pronounced, and when they are weaker. Table 5 examines how the variation in TGR expectations that is explained by firm, industry, and firm-country fixed effects, respectively, changes across the firm's life cycle. In the left column, we find that a firm's industry and country explain only a small portion of TGR expectations in young firms. Specifically, these factors explain only 9% of the variation in very young firms aged 5

[19] These estimates are for the largest sample corresponding to Column (1), Panel A of Table 4. Estimates for the other specifications and subsamples in Table 4 yield similar conclusions.



Figure 5: TGR expectations over firms' life cycle The figure plots analyst terminal growth rates over firms' life cycle. The solid blue line corresponds to Age group fixed effects obtained from the regression: $TGR_{i,j,t} = k + I_{\text{Age group}} + I_{\text{Firm}} + I_{\text{Year}} + \epsilon_{i,j,t}$, for which we add back the regression constant (k) to better illustrate the terminal growth rate level. The pale blue lines correspond to the $I_{\text{Age group}}$ regression estimate 95^{th} confidence interval, for which we add back the regression constant (k) for consistency with the other lines of the figure, and illustration purposes. The variable Firm Age Group captures firm age groups measured in 5-year increments. For example, Firm Age Group = 10 corresponds to firms that are 5 to 10 years old.

years or less. Even 20-25 years after a firm's IPO, industry and country explain only about 20% of TGR forecasts. In the right column, we find that in these young firms, firm fixed effects explain a large share of the variation in TGR expectations that is unaccounted for by industry and country factors, consistent with the evidence in Section 5.1. They account for up to 60% of unexplained variation in very recently IPOed firms, and up to about 50% for firms aged 20-25 years.

These patterns evident in young firms reverse as firms mature. In old firms, the left column of Table 5 shows that a firm's industry and country can explain a much larger amount of the variation in TGR expectations, e.g., 50% for 60-year-old firms and up to 70% and more for 90+-year-old firms. Correspondingly, for mature firms, the incremental explanatory power of firm fixed effects over industry and country strongly diminishes, amounting to, e.g., only up to 14% and 1% for 60-year old and 90+-year-old firms, respectively.

Overall, the evidence above indicates that the explanatory power of firm fixed effects documented in Section 5.1 is particularly pronounced in younger firms and gradually declines with firm age. As firms mature, firm country and industry factors increasingly account for a larger share of TGR expectations.

5.2.2 Dispersion in TGR Expectations Over Firms' Life Cycle

As a bridging step between studying firm and analyst characteristics, we also examine how the within-firm-age *dispersion* of TGR expectations across analysts varies over the firm's life cycle. As Table 6 shows, dispersion in analyst TGR beliefs is largest in young firms. For these, the estimates imply that the average absolute dispersion in TGR expectations in the sample is about 1 pp. The TGR belief dispersion across forecasters declines markedly as firms mature, by as much as 85% for the most mature firms. These patterns are robust to specifications with and without industry-by-year fixed effects as well as a control variable for firm size, and we note that a Bonferroni adjusted test for the joint equality of all the firm age coefficients is rejected at the 1% level in all specifications. These patterns are consistent with analysts exhibiting greater variation in their access to, processing, and interpretation of information related to firm factors compared to industry and country factors that impact their long-term expectations.

5.3 Forecaster Characteristics and Backgrounds

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

Table 7 performs a sequential ANOVA decomposition of estimates of individual analyst fixed effects, which we obtain using a mover design—i.e., analysts moving brokerage houses in our sample, allowing us to separate analyst and employer fixed effects—following the literature in labor economics (Abowd et al., 1999). Jointly, analyst demographics can explain about 30% of the variation in analyst fixed effects, which is a larger amount relative to prior studies on retail investors (cf. Giglio et al., 2021) and households (cf. Malmendier and Nagel,

2011; Bailey et al., 2019; Ben-David et al., 2018; Kuchler and Zafar, 2019; Das et al., 2020; D'Acunto et al., 2023), which typically finds that demographics can explain 10% or even much less of the belief variation.

A plausible explanation for our greater capability to explain heterogeneity in forecaster expectations is the greater diversity in forecaster characteristics and backgrounds within our sample. This sample includes analysts from various countries, in contrast to the more homogeneous samples in prior work that typically includes individuals from a single country and who frequently share additional commonalities (e.g., being affluent U.S. retail investors affiliated with Vanguard in the case of Giglio et al., 2021).

Demographic factors that stand out in Table 7 as being particularly related to the analyst fixed effects in our sample include analyst race, birth year (age), and location (country of residence). This is true both when we examine the full sample of analysts with identified individual fixed effects based on Abowd et al. (1999) in Column (1), when we separate the sample based on whether analysts have both foreign and domestic firm coverage (Column (2)), or when we exclude analysts with domestic coverage (Column (3)).

Table 8 further displays the coefficients on analyst observables when regressed, individually or jointly, on the individual analyst fixed effects. Consistent with Table 7, we observe significant improvements in the explained variation and statistical significance on the coefficients in particular with respect to analyst race characteristics (Column (2)), birth year (Column (4)), and analyst location (broad geographic region in Column (5) and analyst country in Column (6)).[20] The point estimates also reveal the directional relation between analyst demographics and persistent heterogeneity. For example, Asian analysts are more "optimistic" in the sense of having higher TGR expectations, whereas younger analysts born more recently tend to have lower TGR expectations.[21]

[20] In total, analysts in Table 8 are located in 33 different countries, i.e., the country fixed effects in Column (6) are about six times more granular than the country region indicators in Column (5). Whether we include the indicators for White and Asian analysts in Columns (2) to (5) or race fixed effects (which further distinguish non-White and non-Asian analysts) makes little difference empirically.

[21] The large increase in the constant in Columns (4) and (5) arises from the fact that we include birth years without transforming them to be mean-zero.

5.4 Forecasters' Personal Economic Environment and Extrapolative Expectations

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

Table 9 presents the the estimates of how personally, and locally, experienced economic conditions affect forecasters' long-run expectations about firms, exploiting the richness in our data with respect to the location and country of firms and forecasting analysts. In Columns (1) and (3), we find that recent GDP growth experienced by the forecasting analyst in their country of residence strongly predicts the analyst's long-run firm expectations, *controlling for* the recent economic performance in the country in which the firm is located.[22] This pattern points to substantial extrapolation by analysts from local factors when forming their TGR beliefs. This is in particular since the effects of home-country experiences are concentrated in forecasted firms that are located in geographically distant areas (Column (3)). Such a pattern is not predicted by a possible alternative mechanism that is unrelated to extrapolation and in which common shocks to a cluster of neighboring countries affect expectations.

Overall, the evidence in Table 9 extends the existing evidence on local extrapolation (e.g., Kuchler and Zafar, 2019), by documenting strong extrapolative effects across countries and economic environments, present in professional forecasters, and for long-run economic variables.

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

In this section, we open the black box behind analysts' quantitative long-run expectations, and investigate how forecasters' expectation formation process is shaped by differential focus on underlying topics relevant to firm growth prospects. Specifically, we investigate how

[22] As Columns (2) and (4) show, the extrapolation mechanism appears to work through experienced macroeconomic productivity factors, as we find no evidence of extrapolation with respect to locally experienced inflation on long-run expectations.

analysts allocate attention in the qualitative parts of their reports to different topics both on the extensive margin, i.e., their decision to include a given topic in a report, and on the intensive margin, i.e., the intensity with which covered topics are discussed. We then relate differences in topic focus to TGR belief heterogeneity.

5.5.1 Topical Attention Allocation

Topic Focus. We start with a general exploration of the topics discussed in the analyst reports to evaluate the basic properties of topic attention allocation, leveraging the LDA topic model discussed in Section 3.4. An average report is well described by a focus on few topics. Specifically, a report's top 7 topics account, on average, for 84% of the entire text content. Additional topics receive virtually no weight, with each receiving 0.4% pp of the total attention "budget" on average.

Extensive Margin Differences. Two non-mutually exclusive explanations could drive these patterns: (i) professionals *subjectively* focus on a limited number of topics when forming expectations, or (ii) a small number of topics *objectively* characterizes a firm's potential at a given point in time. To differentiate between these explanations, we compare the topics discussed by pairs of analysts evaluating the same firm at the same time. Intuitively, if objective factors are the main driver in analysts' topic focus decision, we would expect that pairs of analysts would mostly agree on the set of discussed topics. By contrast, if analysts' subjective importance of different topics drives the decision process, we would expect differences in the set of topics across analysts.

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

$$\text{Jaccard similarity} = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A and B are the sets of topics discussed in the reports of analysts A and B, $|A \cap B|$ represents the number of topics that overlap in both reports, and $|A \cup B|$ denotes the total number of distinct topics discussed in at least one of the reports. Simply put, Jaccard similarity quantifies the overlap between the sets of topics included in both reports, ranging

from 0 (no overlap) to 1 (full overlap). In our baseline analysis, a topic is assumed to be *covered* when it accounts for at least 1% of the qualitative discussion.

For the average analyst pair, 43% of topics covered by at least one analyst of the pair are included in both reports with an interquartile range of 17 pp (33% to 50%).[23] To make this magnitude more concrete, a Jaccard score of 40% could represent a case where analyst A focuses on topics 1, 2, 3, 4, and analyst B covers topics 1, 2, and 5. The Jaccard similarity scores indicate that analysts tend to agree on some key topics that must be included in reports, but there remain substantial differences across analysts regarding——or, put differently, person-specific factors determining——the full set of topics that are covered in a report.

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

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

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

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

For the average analyst pair, we obtain a Cosine similarity of 0.79 (with an interquartile range between 0.72 and 0.91), i.e., an average dissimilarity in topic weights of approximately 0.2. As a reference benchmark, Appendix Table B.1 provides the Cosine similarity distributions

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

for overlapping topics between firms depending on whether these operate in the same or different industries. The observed average Cosine similarity of 0.79 across forecasters, while high at first glance, is only slightly higher than the average Cosine similarity of firms operating in different industries (0.74). This implies only a small incremental "convergence" of analysts forecasting the same firm at the same time in terms of chosen topic weights, relative to a cross-industry benchmark.

Overall, the results presented in this section suggest that the LDA model estimation and analysis uncover meaningful differences in topic attention allocation across forecasters, both at the extensive and intensive margin.[24]

5.5.2 Topic Attention Differences and Variation in TGR Expectations

Having established the differences in analysts' topic focus, we examine the extent to which these differences predict the heterogeneity in analysts' long-term TGR expectations, thus connecting quantitative and qualitative aspects of the belief formation process.

We study both the extensive margin of topic focus, using the Jaccard similarity, and the intensive margin of topic weights, employing the Cosine similarity of overlapping topics. Simply put, we ask whether differences in TGR expectations for a pair of analysts are mainly driven by the fact that they cover different topics, or because they discuss the same topics differently.

Working with pairwise measures allows us to leverage the particular nature of our setting, the fact that we often have multiple professionals evaluating the same firm at the same time. It also allows us to explore new dimensions of attention–extensive and intensive margin–instead of the topic themselves. Finally, it offers a novel and alternative strategy to address the LDA output dimensionality problem by considering topics that are not systematically covered by all analysts or for all firms in a given subgroup, contrasting with other strategies that have been previously used in the literature, such as using Lasso regressions (Bybee et al., 2023a).

[24] We stress that these patterns are not driven by differences in report length between pairs of analysts. Both Jaccard and Cosine similarity capture differences in topic coverage as a fraction of the total report, offering a unit-less metric, instead of counting words. Further, we obtain the same average magnitude when restraining the sample to cases where the length across reports differs by less than 1% (Jaccard score of 43% and Cosine score of 80%, respectively).

Table 10 presents the results. The dependent variable is the pairwise difference in TGR beliefs across analysts:

$$TGR_{A,B,i,t} = \frac{(TGR_{A,i,t} - TGR_{B,i,t})^2}{|(TGR_{A,i,t} + TGR_{B,i,t})/2|} \tag{3}$$

where A and B refer to the analysts in the pair, i refers to the firm covered in the report, and t the year of the reports. Intuitively, large values of $TGR_{A,B,i,t}$ suggest that there are large relative differences in the TGR beliefs of analysts. Columns (1) to (3) show the extensive margin results, where the variable of interest is the Jaccard similarity, and Columns (4) to (6) present the intensive margin results, where the variable of interest is the Cosine similarity. Across all columns, the coefficients of interest on the Jaccard and Cosine similarity measures are negative, consistent with the notion that more similar discussions in equity reports negatively predict the disagreement about and heterogeneity in TGR expectations.

Additionally, we estimate robust differences in the strength of the association between quantitative beliefs and qualitative topic discussions, depending on the extensive versus intensive margin of topic overlap. The extensive-margin Jaccard regression coefficients are, while negative, notstatistically significant (p-value between 0.21 and 0.34), so in fact we cannot reject a null relationship. By contrast, the intenstive-margin Cosine point estimates are large (-0.60) and highly significant at 5% across all specifications, with Column (4) including no fixed effects, Column (5) using year and industry (2 digits NAICS code) fixed effects, and Column (6) utilizizing year interacted with industry fixed effects.[25] We also note that because both our dependent variable and variables of interest are measured as differences at the firm and year level, the analysis is naturally conducted at the within-firm-year level. Lastly, Column (7) includes both the Jaccard and Cosine measures and validates our conclusion from the univariate regressions in Columns (1) to (6).

Overall, these results indicate that differences in long-term expectations among professionals mainly stem from prioritizing overlapping topics differently, rather than discussing different topics altogether.

[25] The fixed effects account for broad trends that would occur at the industry level over time.

5.5.3 Understanding Topic Weight Heterogeneity

In a final set of analyses, we seek to further understand the details and determinants of differences in the intensive-margin topic focus of forecasters, and validate that the topic focus results from Table 10 are indeed driven by *economically intuitive* differences in forecasters' topic focus and attention.

To this end, we first break down the Cosine similarity measure into five broad categories: (i) macroeconomics, (ii) industry, (iii) geography, (iv) operations, and (v) valuation, to identify which category or categories of topics relate the most to differences in long-term expectations.[26] We then reestimate three versions of the Cosine similarityscore, by separately (and exclusively) using the topics associated with the macroeconomic, geographic, or industry category. In Table 11, we find a negative and statistically significant relation between each of these category-specific Cosine similarity measures and the variation in long-run beliefs. Additionally, and consistent with our evidence on local extrapolation, differences in expectations appear to be shaped the most by differences in the attention allocation intensity for geographic and macroeconomic topics, given their larger (in absolute value) regression coefficients. For example, the results in Table 11 imply that a, say, Argentinian analyst is likely to discuss in greater detail a firm's operation in South America and expand more on the potential effect of inflation compared to, say, an American analyst.

Finally, to further explore the role of geographic location, we investigate whether the distance between forecasters is related to the way they cover topics in their reports at the intenstive margin. In Columns (1) to (3) of Table 12, we find that analysts that are located further apart are more likely to discuss the same topics in a different way. To tighten the connection to our evidence on local extrapolation from Table 9, Columns (4) to (6) focus on analysts' geographic coverage. We find that analysts that are located far away from each other tend to allocate topic attention differently when discussing firms' geographic activities, suggesting that, say, Argentinian and Brazilian analysts are more likely to discuss a firm's geographic footprint similarly than are, say, American and Argentinian analysts.

Overall, local conditions and experiences appear to influence professionals' coverage

[26] Appendix Table B.2 provides a list of our LDA-based topics with their associated categories.

intensity, serving as one underlying mechanism for extrapolation.

6 Conclusion

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

We establish a number of novel facts and findings. TGR expectations convey distinct economic insights compared to shorter-term forecasts such as IBES long-term growth (LTG). They robustly forecast realized long-run firm growth and, in line with firm life cycle models, decline with firm age. There is persistent heterogeneity in TGR expectations across forecasters, particularly for less mature firms where country and industry factors do not predominate, and forecaster demographics play a substantial role in explaining this variation. Forecasters extrapolate from economic conditions in their country of residence when forming long-run expectations about firms in other, distant countries. Finally, varying topic foci in forecasters' qualitative discussions emerge as one underlying mechanism for long-run belief heterogeneity and extrapolation.

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

References

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

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

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.

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.

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. LaPorta, and A. Shleifer. 2023. Belief Overreaction and Stock Market Puzzles. *American Economic Review* pp. 1102–1130.

Bouchaud, J.-P., P. Krüger, A. Landier, and D. Thesmar. 2019. Sticky expectations and the profitability anomaly. *The Journal of Finance* 74:639–674.

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

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

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

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

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

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., and S. Myers. 2023. Which Subjective Expectations Explain Asset Prices? .

Decaire, P. H. 2023. Discount Rate Uncertainty and Capital Investment. *Working paper* .

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.

Gabaix, X. 2014. A Sparsity-Based Model of Bounded Rationality. *Quarterly Journal of Economics* 129:1661–1710.

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* .

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* .

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.

Hommel, N., A. Landier, and D. Thesmar. 2023. Corporate Valuation: An Empirical Comparison of Discounting Methods. *Working paper* .

Huang, S., H. Tan, X. Wang, and C. Yu. 2023. Valuation Uncertainty and Analysts' Use of DCF Models. *Review of Accounting Studies* 28:827–861.

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

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

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

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

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

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

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

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

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

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

Panel A: Firms and Coverage						
		No. Firm			No. Obs	
Terminal growth rate sample		8,010			41,429	
Equity Analyst					5,304	
Equity Analyst (LinkedIn Matched)					1,667	
Brokerage house					41	
Panel B: Equity Reports	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
No. foreign reports	0.31	0.00	0.00	1.00	0.46	41,129
No. reports by country	845.49	11.00	240.00	581.00	2,140.80	49
Analyst terminal growth rate $(TGR)_{i,j,t}$(%)	2.22	1.50	2.00	3.00	1.31	41,429
$TGR_{i,j,t}$(%) (IBES Overlap)	2.21	1.50	2.00	3.00	1.29	25,615
IBES $LTG_{i,j,t}$(%)	13.03	5.50	10.30	17.00	22.48	25,615
Correlation(IBES LTG, TGR)	0.08^{***}					
Panel C: Analyst Information	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Female	0.18	0.00	0.00	0.00	0.38	1,667
Asian	0.18	0.00	0.00	0.00	0.38	1,667
Black	0.03	0.00	0.00	0.00	0.18	1,667
Caucasian	0.75	1.00	1.00	1.00	0.43	1,667
Hispanic	0.03	0.00	0.00	0.00	0.18	1,667
Graduate degree (Masters and MBAs)	0.44	0.00	0.00	1.00	0.50	1,667
Birth year	1977.77	1971.00	1978.00	1985.00	9.64	1,020
No. Reports by Analyst	7.92	1.00	3.00	7.00	16.41	5,304
No. analyst by country	120.14	9.00	48.00	91.50	259.85	44
No. analyst by brokerage house	142.15	24.00	46.00	180.00	186.16	41
Panel D: Economic Data Series—All sample	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Firm country						
Real GDP $growth_{i,j,t}$ (%)	2.55	1.55	2.29	3.74	3.36	38,370
$Inflation_{i,j,t}$ (%)	2.57	1.26	2.01	3.16	2.50	38,247
10-year Treasury $yield_{i,j,t}$ (%)	3.15	1.68	2.91	4.10	2.30	40,000
Panel E: Economic Data Series—Foreign reports	Mean	25^{th} Pct.	Median	75^{th} Pct.	Std. Dev.	No. Obs.
Analyst country						
Real GDP $growth_{i,j,t}$ (%)	2.01	1.45	2.29	3.12	3.70	12,296
$Inflation_{i,j,t}$ (%)	2.49	1.45	2.27	2.85	1.93	12,285
10-year Treasury $yield_{i,j,t}$ (%)	3.02	1.51	2.92	4.14	2.10	12,828
Firm country						
Real GDP $growth_{i,j,t}$ (%)	2.11	0.96	2.21	3.82	3.58	11,802
$Inflation_{i,j,t}$ (%)	2.43	0.90	1.83	3.06	2.75	11,686
10-year Treasury $yield_{i,j,t}$ (%)	3.02	0.97	2.75	4.11	2.76	11,830

Table 2: Realized long-term earnings growth rates versus analysts' expected TGRs In Panel A, this table decomposes ex-post realized earnings growth at different horizons based on analysts' terminal growth rates. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023. The dependent variables in Columns 1 to 3 are the geometric average of firms' earnings annual growth rates experienced by firm i for the 5-year periods $t+n$ to $t+n+4$, for different values of n. *Terminal growth rate*$_{i,j,t}$ is the terminal growth rate expectation by the analyst, measured in percentage points. In Panel B, we replicate the analysis from Panel A, but we average the geometric averages in each year across observations. The unit of observation is at the forecast year t level. 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: Firm level	$\Pi_{t+1}^{t+5}(1+g_{i,t})^{(1/5)}-1$	$\Pi_{t+4}^{t+8}(1+g_{i,t})^{(1/5)}-1$	$\Pi_{t+4}^{t+13}(1+g_{i,t})^{(1/10)}-1$
	(1)	(2)	(3)
Terminal growth rate (%)$_{i,j,t}$	1.02***	0.70***	0.53***
	(0.12)	(0.12)	(0.10)
Constant	4.42***	3.38***	2.93***
	(0.31)	(0.30)	(0.29)
Observations	20,812	17,116	9,367
F Statistics	70.03	36.66	28.58
R^2	0.01	0.00	0.01
Panel B: Sample average	$\frac{\sum_{j=1}^{J}\Pi_{t+1}^{t+5}(1+g_{j,t})^{(1/5)}-1}{J}$	$\frac{\sum_{j=1}^{J}\Pi_{t+4}^{t+8}(1+g_{j,t})^{(1/5)}-1}{J}$	$\frac{\sum_{j=1}^{J}\Pi_{t+4}^{t+13}(1+g_{j,t})^{(1/10)}-1}{J}$
	(1)	(2)	(3)
Average terminal growth rate (%)$_t$	4.18**	4.19***	1.91***
	(1.73)	(1.00)	(0.22)
Constant	-2.89	-4.49*	-0.40
	(3.90)	(2.49)	(0.51)
Observations	16	14	8
F Statistics	5.86	17.62	76.65
R^2	0.34	0.44	0.60

Table 3: TGR, firm age, and default risk This table presents the relation between firm age, default risk, and analyst terminal growth rates. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023. The dependent variable, *Terminal growth rate*$_{i,j,t}$, is the terminal growth rate expectation by the analyst, measured in percentage points. Default risk variables are indicator variables equal to 1 if the firm probability of default is above their respective sample median, and 0 otherwise. The firm probability of default is estimated using the naive Merton distance-to-default model described in Bharath and Shumway (2008), using a 10- and 20-year horizon ($T = 10$ or 20), respectively. Age is the natural logarithm of firm age measured from the IPO year. 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%.

Dependent variable	Terminal growth rate$_{i,j,t}$				
	(1)	(2)	(3)	(4)	(5)
Default risk (10-year horizon)$_{i,j,t}$	-0.16***			-0.18***	
	(0.03)			(0.04)	
Default risk (20-year horizon)$_{i,j,t}$		-0.18***			-0.21***
		(0.03)			(0.04)
ln(Firm post-IPO age$_{i,j,t}$)			-0.14***	-0.17***	-0.17***
			(0.02)	(0.03)	(0.03)
Firm Country*Industry*Year FE	Yes	Yes	Yes	Yes	Yes
Observations	26,172	26,172	22,124	14,745	14,745
F Statistics	39.93	47.97	41.68	26.82	31.96
R^2	0.33	0.33	0.33	0.36	0.36

Table 4: ANOVA variance decomposition This table presents a variance decomposition of analyst terminal growth rates using ANOVA. For this analysis, we exclude from the sample firms, analysts, and brokerage houses with fewer than 2 observations in total, as for these observations, their respective fixed effects fully explain the associated variation in terminal growth rates. We also exclude brokerage houses that are not associated with equity analysts that switched between employers during the sample period (2000–2023), since it is not possible to distinguish between the effect of those brokerage houses and their analysts (Abowd et al., 1999). Panel A includes this entire subsample. Panel B excludes firms with fewer than 10 observations. To account for the fact that the sample used to conduct the analysis is unbalanced, as we do not observe a terminal growth rate estimate for every analyst-firm-year pair during the entire sample period, we estimate the sequential partial sum of squares. The adjusted partial R^2 is defined as: $1-(1-(\frac{Pss_i}{TSS})*\frac{N-1}{N-df_i-1})$, where i denotes a specific set of fixed effects (brokerage house, analyst, or firm). Variable definitions appear in Appendix A.

	Panel A: Anova			
	All sample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$
Brokerage house FE				
Share of total R^2 (%)	2	3	4	12
Degrees of Freedom	34	34	32	25
Adjusted Partial R^2	0.01	0.02	0.02	0.08
Analyst FE				
Share of total R^2 (%)	71	65	61	48
Degrees of Freedom	3,500	1,018	384	94
Adjusted Partial R^2	0.42	0.42	0.41	0.34
Firm FE				
Share of total R^2 (%)	27	32	35	40
Degrees of Freedom	5,007	4,434	3,335	1,902
Adjusted Partial R^2	0.04	0.06	0.06	0.09
Total R^2	0.67	0.67	0.69	0.72
Observations	35,719	26,565	17,733	8,504
	Panel B: Anova—Obs per firm ≥ 10			
	All subsample	$\frac{Reports}{Analyst} \geq 10$	$\frac{Reports}{Analyst} \geq 25$	$\frac{Reports}{Analyst} \geq 50$
Brokerage house FE				
Share of total R^2 (%)	3	4	6	15
Degrees of Freedom	34	34	32	25
Adjusted Partial R^2	0.02	0.02	0.04	0.10
Analyst FE				
Share of total R^2 (%)	78	73	67	52
Degrees of Freedom	2,694	944	373	103
Adjusted Partial R^2	0.42	0.42	0.41	0.34
Firm FE				
Share of total R^2 (%)	18	23	27	33
Degrees of Freedom	1,187	1,173	1,105	767
Adjusted Partial R^2	0.06	0.08	0.08	0.09
Total R^2	0.63	0.62	0.64	0.68
Observations	21,528	16,651	11,480	5,456

Table 5: TGR expectations over firms' life cycle This table evaluates the share of the second-stage dependent variable explained by the right-hand-side explanatory variables over firm life cycle. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares.

Stage 1:	None	$TGR_{i,j,t} = I_{\text{Firm country}} + I_{\text{Industry}} + \epsilon_{i,j,t}$
Stage 2:	$TGR_{i,j,t} = I_{\text{Firm country}} + I_{\text{Industry}} + \epsilon_{i,j,t}$	$\hat{\epsilon}_{i,j,t} = I_{\text{Firm}} + \upsilon_{i,j,t}$
	Stage 2 R^2	**Stage 2** R^2
0 <= Firm age < 5	0.09	0.60
5 <= Firm age < 10	0.09	0.55
10 <= Firm age < 15	0.10	0.53
15 <= Firm age < 20	0.14	0.52
20 <= Firm age < 25	0.18	0.54
25 <= Firm age < 30	0.24	0.51
30 <= Firm age < 35	0.18	0.46
35 <= Firm age < 40	0.19	0.62
40 <= Firm age < 45	0.20	0.43
45 <= Firm age < 50	0.38	0.37
50 <= Firm age < 55	0.32	0.41
55 <= Firm age < 60	0.25	0.23
60 <= Firm age < 65	0.56	0.14
65 <= Firm age < 70	0.57	0.34
70 <= Firm age < 75	0.64	0.11
75 <= Firm age < 80	0.45	0.16
80 <= Firm age < 85	0.60	0.16
85 <= Firm age < 90	0.28	0.02
90 <= Firm age < 95	0.72	0.01
95 <= Firm age < 100	0.84	0.00

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

Terminal Growth Rate dispersion	$(TGR_{i,j,t} - \overline{TGR}_{k,t}^{\text{Age Group}})^2/\overline{TGR}_{k,t}$		
	(1)	(2)	(3)
5 <= Firm age < $10_{i,j,t}$	-0.01	-0.01	-0.03
	(0.03)	(0.02)	(0.03)
10 <= Firm age < $15_{i,j,t}$	-0.04	-0.03	-0.05*
	(0.03)	(0.03)	(0.03)
15 <= Firm age < $20_{i,j,t}$	-0.08**	-0.08***	-0.10***
	(0.03)	(0.03)	(0.03)
20 <= Firm age < $25_{i,j,t}$	-0.06*	-0.07**	-0.10***
	(0.03)	(0.03)	(0.04)
25 <= Firm age < $30_{i,j,t}$	-0.04	-0.06	-0.09**
	(0.04)	(0.04)	(0.04)
30 <= Firm age < $35_{i,j,t}$	-0.17***	-0.20***	-0.26***
	(0.04)	(0.04)	(0.04)
35 <= Firm age < $40_{i,j,t}$	-0.27***	-0.30***	-0.38***
	(0.04)	(0.04)	(0.04)
40 <= Firm age < $45_{i,j,t}$	-0.32***	-0.33***	-0.38***
	(0.05)	(0.05)	(0.06)
45 <= Firm age < $50_{i,j,t}$	-0.42***	-0.40***	-0.46***
	(0.05)	(0.05)	(0.05)
50 <= Firm age < $55_{i,j,t}$	-0.37***	-0.37***	-0.39***
	(0.06)	(0.07)	(0.07)
55 <= Firm age < $60_{i,j,t}$	-0.45***	-0.47***	-0.54***
	(0.05)	(0.06)	(0.07)
60 <= Firm age < $65_{i,j,t}$	-0.26***	-0.34***	-0.41***
	(0.09)	(0.09)	(0.10)
65 <= Firm age < $70_{i,j,t}$	-0.54***	-0.52***	-0.58***
	(0.06)	(0.05)	(0.05)
70 <= Firm age < $75_{i,j,t}$	-0.28***	-0.35***	-0.45***
	(0.10)	(0.09)	(0.10)
75 <= Firm age < $80_{i,j,t}$	-0.29***	-0.36***	-0.24
	(0.09)	(0.11)	(0.15)
80 <= Firm age < $85_{i,j,t}$	-0.46***	-0.49***	-0.54***
	(0.09)	(0.13)	(0.15)
85 <= Firm age < $90_{i,j,t}$	-0.54***	-0.43***	-0.46***
	(0.07)	(0.09)	(0.09)
90 <= Firm age < $95_{i,j,t}$	-0.30*	-0.36***	-0.41***
	(0.16)	(0.12)	(0.13)
95 <= Firm age < $100_{i,j,t}$	-0.61***	-0.60***	-0.64***
	(0.03)	(0.05)	(0.06)
Firm $\text{Size}_{i,t}$			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 the same (Bonferroni p-value)	0.00	0.00	0.00
Observations	24,280	24,255	21,295
F Statistics	37.53	19.07	16.57
R^2	0.01	0.08	0.08

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

Panel A: Fixed effects	Terminal growth rate$_{i,j,t}$		
	All Sample	Has both foreign and domestic coverage	Exclude domestic coverage
	(1)	(2)	(3)
Constant	2.17***	2.17***	2.14***
	(0.00)	(0.00)	(0.00)
Firm FE	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes
Analyst FE	Yes	Yes	Yes
Observations	6,668	2,759	1,355
R^2	0.69	0.71	0.72
Panel B: ANOVA decomposition	**Analyst FE**		
	All Sample	Has both foreign and domestic coverage	Exclude domestic coverage
	(1)	(2)	(3)
Gender FE			
Share of total R^2 (%)	0	1	0
Degrees of Freedom	1	1	1
Adjusted Partial R^2	0.00	0.00	-0.01
Race FE			
Share of total R^2 (%)	9	4	3
Degrees of Freedom	3	3	3
Adjusted Partial R^2	0.02	0.01	0.02
Education FE			
Share of total R^2 (%)	15	1	1
Degrees of Freedom	1	1	1
Adjusted Partial R^2	0.04	-0.01	-0.01
Birthyear FE			
Share of total R^2 (%)	19	46	61
Degrees of Freedom	48	44	36
Adjusted Partial R^2	-0.04	0.00	-0.03
Analyst country FE			
Share of total R^2 (%)	58	47	35
Degrees of Freedom	31	28	15
Adjusted Partial R^2	0.12	0.06	0.08
Total R^2	0.31	0.29	0.43
Observations	479	318	153

Table 8: Analyst fixed effects by analyst demographics This table explores the sign and magnitude associated with observable analyst characteristics in explaining analyst fixed effects. For this analysis, we exclude from the sample firms, analysts, and brokerage houses with fewer than 2 observations in total, as for these observations, their respective fixed effects fully explain the associated variation in terminal growth rates. We also exclude brokerage houses that are not associated with equity analysts that switched between employers during the sample period (2000–2023), since it is not possible to distinguish between the effect of those brokerage houses and their analysts (Abowd et al., 1999). The unit of observation is at the analyst j level. Female is a binary variable equal to 1 if the analyst is female and 0 otherwise. White is a binary variable equal to 1 if the analyst is white, and 0 otherwise. Asian is a binary variable equal to 1 if the analyst is Asian, and 0 otherwise. Graduate degree is a binary variable equal to 1 if the analyst has a graduate degree, and 0 otherwise. North America, South America, Europe, Asia, and Oceania are binary variables equal to 1 if the analyst is located on the given continent, and 0 otherwise. The omitted category is Africa. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent. Variable definitions appear in Appendix A. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable	Analyst FE					
	(1)	(2)	(3)	(4)	(5)	(6)
I_j^{Female}	-0.02	-0.11	-0.11	-0.07	-0.03	
	(0.19)	(0.20)	(0.20)	(0.20)	(0.19)	
I_j^{White}		-0.30	-0.30	-0.30	-0.21	
		(0.26)	(0.26)	(0.25)	(0.26)	
I_j^{Asian}		0.47	0.47	0.54*	-0.08	
		(0.31)	(0.31)	(0.31)	(0.30)	
$I_j^{\text{Graduate degree}}$			0.01	-0.03	-0.02	
			(0.12)	(0.12)	(0.12)	
Birth year$_j$				-0.02***	-0.01**	
				(0.01)	(0.01)	
$I_j^{\text{North America}}$					-1.15***	
					(0.43)	
$I_j^{\text{South America}}$					-1.74*	
					(0.99)	
I_j^{Europe}					-1.42***	
					(0.41)	
I_j^{Asia}					-0.16	
					(0.50)	
I_j^{Oceania}					-1.18***	
					(0.42)	
Constant	0.10	0.26	0.25	35.48***	30.19**	0.10*
	(0.07)	(0.25)	(0.26)	(12.66)	(12.63)	(0.06)
Gender FE	No	No	No	No	No	Yes
Race FE	No	No	No	No	No	Yes
Education FE	No	No	No	No	No	Yes
Birthyear FE	No	No	No	No	No	Yes
Analyst country FE	No	No	No	No	No	Yes
Observations	486	486	486	486	486	477
F Statistics	0.01	4.82	3.62	4.40	5.34	.
R^2	0.00	0.04	0.04	0.06	0.11	0.27

Table 9: Analyst–firm distance, analyst local economic conditions, and extrapolation This table explores the effect of analyst local economic conditions on their terminal growth rate expectations. The unit of observation is at the firm i, analyst j, and forecast year t levels. The sample period is 2000–2023. The dependent variable, *Terminal growth rate*$_{i,j,t}$, is equal to the terminal growth rate expectation by the analyst, measured in percentage points. The 10-year average real GDP growth series, for both the firm and the analyst's country, are taken from the World Bank data series. They are measured as the trailing 10-year average. The 10-year average inflation rates, for both the firm and the analyst's country, are constructed from data from the World Bank. They are measured as the trailing 10-year average. *Far* is a binary variable equal to 1 if the geographic distance between the analyst's country and the firm's country is above the sample median, and 0 otherwise. We measure the distance between countries as the shortest distance between the two countries' capitals. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the firm level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable	Terminal growth rate$_{i,j,t}$			
	(1)	(2)	(3)	(4)
10-year average real GDP growth (Firm)$_{i,j,t}$	0.09***		0.09***	
	(0.02)		(0.02)	
10-year average real GDP growth (Analyst)$_{i,j,t}$	0.09***		-0.04	
	(0.03)		(0.05)	
10-year average inflation rate (Firm)$_{i,j,t}$		0.00		0.00
		(0.00)		(0.00)
10-year average inflation rate (Analyst)$_{i,j,t}$		-0.00		-0.00
		(0.00)		(0.00)
10-year average real GDP growth (Analyst)$_{i,j,t} * I^{far}_{i,j,t}$			0.15***	
			(0.05)	
10-year average inflation rate (Analyst)$_{i,j,t} * I^{far}_{i,j,t}$				0.02
				(0.03)
Analyst country*firm country FE	Yes	Yes	Yes	Yes
Observations	11,675	11,575	11,675	11,575
F Statistics	29.82	0.02	21.26	0.12
R^2	0.15	0.14	0.15	0.14

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

Dependent variable	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

Table 11: Variation in TGR Expectations and Category-Specific Topic Weights This table explores the relation between differences in textual discussion and disagreement in terminal growth rate expectations between pairs of analysts. The unit of observation is at the analyst a, analyst b, firm i, and forecast year t levels. The sample period is 2000–2023. The dependent variable, $Disagreement_{a,b,i,t}$, is equal to the difference between analyst A and analyst B TGR estimates squared and scaled by the absolute value of their sum: $(TGR_{a,i,t} - TGR_{b,i,t})^2 / |TGR_{a,i,t} + TGR_{b,i,t}|$. The variables of interest are Cosine scores respectively measured from the subset of topics that only relate to *macroeconomic*, *industry*, or *geographic* topics. Topic classifications are shown in Internet Appendix Table B.2. 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)
$Cosine^{Macro}_{a,b,i,t}$	-1.32**	-1.29**				
	(0.61)	(0.60)				
$Cosine^{Industry}_{a,b,i,t}$			-0.97**	-0.90*		
			(0.46)	(0.48)		
$Cosine^{Geography}_{a,b,i,t}$					-2.16***	-1.65**
					(0.80)	(0.79)
Industry*Year FE	No	Yes	No	Yes	No	Yes
Observations	11,644	11,578	17,688	17,637	11,053	11,008
F Statistics	4.74	4.54	4.40	3.55	7.27	4.32
R^2	0.00	0.07	0.00	0.06	0.00	0.09

Table 12: Geographic Distance Between Analysts and Differences in Topic Attention Allocation
This table explores the relation between geoggraphic distance between pairs of analysts and differences in their intensive-margin topic focus. The unit of observation is at the analyst a, analyst b, firm i, and forecast year t levels. The sample period is 2000–2023. The are two dependent variables. $Cosine_{a,b,i,t}$, denotes the Cosine similarity score across all the topics that overlap in both articles. $Cosine^{Geography}_{a,b,i,t}$, denotes the Cosine similarity score across the geography topics that overlap in both articles. Topic classifications are shown in Internet Appendix Table B.2. $I^{far}_{a,b,t}$ is a binary variable equal to 1 if the geographic distance between the analyst's A country and analyst's B country is above the sample median, and 0 otherwise. We measure the distance between countries as the shortest distance between the two countries' capitals. Variable definitions appear in Appendix A. The regressions are estimated using ordinary least squares, and the standard errors (in parentheses) are heteroskedastic-consistent and clustered at the analyst pair level. Significance levels are shown as follows: * = 10%, ** = 5%, *** = 1%.

Dependent variable	$\text{Cosine}_{a,b,i,t}$			$\text{Cosine}^{\text{Geography}}_{a,b,i,t}$		
	(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

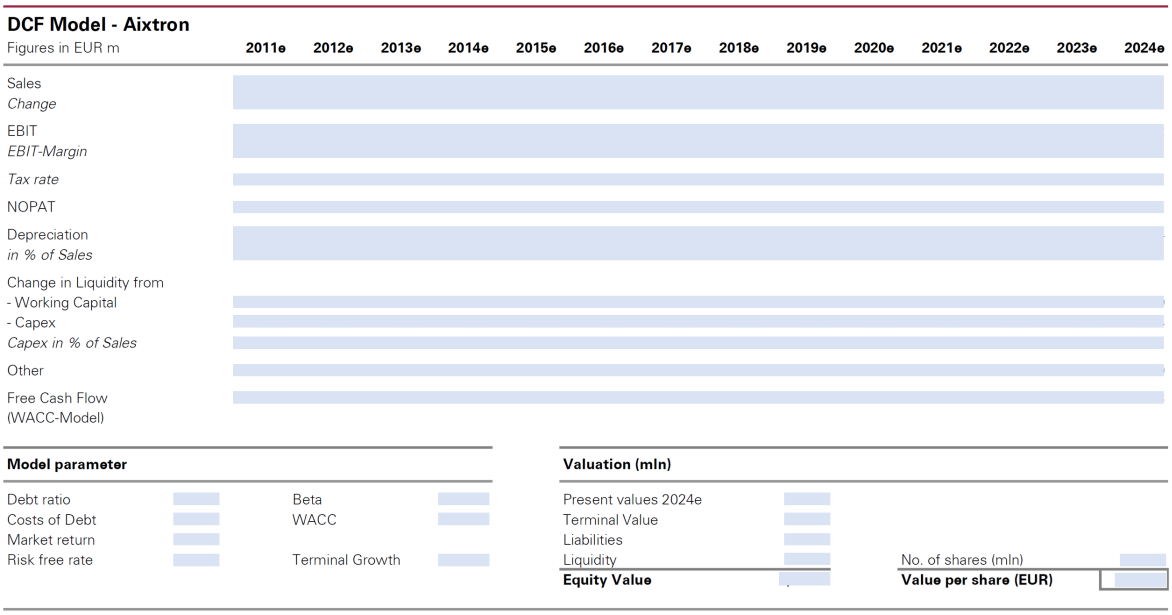
Table A.1: Variable Definitions

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

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

Appendix B Supplementary Figures and Tables

Appendix Figures



DCF Model - Aixtron

Figures in EUR m	2011e	2012e	2013e	2014e	2015e	2016e	2017e	2018e	2019e	2020e	2021e	2022e	2023e	2024e
Sales														
Change														
EBIT														
EBIT-Margin														
Tax rate														
NOPAT														
Depreciation														
in % of Sales														
Change in Liquidity from														
- Working Capital														
- Capex														
Capex in % of Sales														
Other														
Free Cash Flow (WACC-Model)														

Model parameter			
Debt ratio		Beta	
Costs of Debt		WACC	
Market return			
Risk free rate		Terminal Growth	

Valuation (mln)			
Present values 2024e			
Terminal Value			
Liabilities			
Liquidity		No. of shares (mln)	
Equity Value		**Value per share (EUR)**	

Figure B.1: Example of Complete Equity Report DCF This figure shows a representative example of discounted cash flow models when analysts supplement their recommendations with valuation models. This figure is taken from the Aixtron (Ticker = AIXGn) equity report, published by Warburg Research on October 27, 2011. *For copyright reasons, we redacted any information provided in the table.*

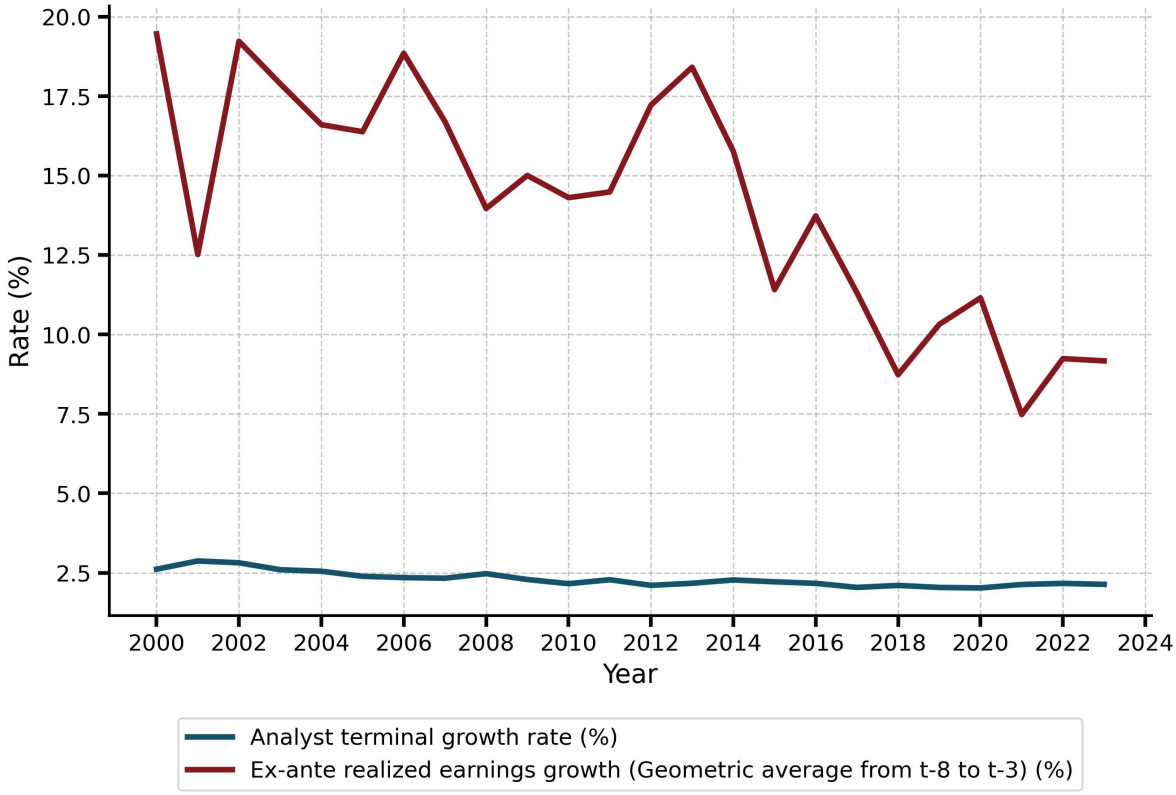


Figure B.2: TGR expectations versus historical returns This figure compares the time trends of historically realized earnings growth at the $t-8$ to $t-3$ horizon with analyst terminal growth rates. Realized earnings growth corresponds to the geometric average realized from $t-8$ to $t-3$, averaged by year (t) in the sample. Terminal growth rate is the average terminal growth rate by year. The x-axis denotes years. The y-axis corresponds to rates, in percentages (%).

Appendix Tables

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

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

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

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

Continued on next page.

Continued from previous page.

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

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

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