The Wisdom of the Crowd? Firm Expectations, Disagreement, and Real GDP Forecast Accuracy*

Brent Meyer†
Federal Reserve Bank of Atlanta

Emil Mihaylov ‡
Federal Reserve Bank of Atlanta

Michael A. Navarrete§
Federal Reserve Bank of Atlanta

Daniel Weitz¶
Federal Reserve Bank of Atlanta

October 5, 2026

Abstract

We compare probabilistic real GDP growth forecasts from the CFO Survey with those from the Survey of Professional Forecasters (SPF). Both surveys elicit full probability distributions, so in addition to forecast accuracy, we can measure subjective uncertainty within forecasters and disagreement across them. CFO forecasts are substantially more dispersed than SPF forecasts: cross-sectional disagreement among CFOs is roughly three times as large as among professional forecasters. At the individual level, CFO respondents exhibit larger absolute errors and higher forecast-error variance. Yet this greater individual-level noise does not translate into worse consensus forecasts. Aggregated CFO forecasts track SPF forecasts closely and achieve statistically similar consensus accuracy. Decompositions of ex-ante forecast variance and ex-post mean squared error show that CFO disagreement reflects heterogeneous beliefs and that much of their individual forecast error is idiosyncratic, allowing aggregation to reduce noise. A within-quarter randomization exercise quantifies this wisdom-of-crowd effect directly: the consensus forecast error declines with crowd size with an elasticity of -0.17, evidence of genuine aggregation gains. Dispersed firm-level information can thus complement professional macroeconomic forecasts, though consensus accuracy is ultimately bounded by shared errors rather than crowd size.

JEL Code: C53, D84, E37
Keywords: firm expectations; professional forecasters; GDP forecasting; wisdom of crowd; CFO Survey; Survey of Professional Forecasters

*We thank Chen Lian for helpful comments. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta, the Federal Reserve System, the Board of Governors, or their staff. All errors are our own.

†Email: brent.meyer@atl.frb.org. Address: 1000 Peachtree Street NE, Atlanta, GA 30309.
‡Email: emil.mihaylov@atl.frb.org. Address: 1000 Peachtree Street NE, Atlanta, GA 30309.
§Email: michael.navarrete@atl.frb.org. Address: 1000 Peachtree Street NE, Atlanta, GA 30309.
¶Email: daniel.weitz@atl.frb.org. Address: 1000 Peachtree Street NE, Atlanta, GA 30309.

"With most things, the average is mediocrity. With decision making, it's often excellence."

— James Surowiecki, *The Wisdom of Crowds*, Part I, Ch. 1

1 Introduction

Survey data increasingly provide a way to study not only how agents form expectations, but also whether those expectations can be aggregated to forecast macroeconomic outcomes. In this paper, we ask whether business leaders can forecast aggregate real GDP growth as well as, or better than, professional forecasters. The question is motivated by a simple wisdom-of-crowd idea: even if individual firm forecasts are noisy, aggregating across many firms may average out idiosyncratic errors and produce accurate forecasts of aggregate outcomes. We study this possibility using the CFO Survey and compare its forecasts to those from the Survey of Professional Forecasters (SPF), a natural benchmark for professional macroeconomic forecasting. Beyond this horse race, the two surveys may be complements rather than substitutes. Business leaders draw on information that professional forecasters do not observe, including knowledge of their own operations, industries, customers, and supply chains, and the two surveys elicit forecasts at different horizons. Whether firm expectations carry information that professional forecasts do not already contain is ultimately what determines their value to policymakers and forecasters.

A distinctive feature of our setting is that both surveys elicit probability distributions over real GDP growth. This allows us to compare the CFO Survey and the SPF on more than mean forecast accuracy. We can measure disagreement across forecasters and subjective uncertainty within forecasters, and show that disagreement raises individual forecast errors but averages out in the consensus.. Because the two surveys ask closely related probabilistic questions about real GDP growth, we can conduct a comparison of both forecast accuracy and uncertainty at the individual and aggregate levels. In addition, the CFO Survey's large and time-varying cross-section allows us to examine whether forecast accuracy improves as the number of forecasters increases—a direct test of the aggregation mechanism at the heart of the wisdom-of-crowd idea.

Our analysis is organized around three questions. First, are CFOs systematically less accurate than SPF forecasters when predicting real GDP growth, and does the answer change after aggregating

individual forecasts? Second, does the CFO Survey exhibit the signatures of a wisdom-of-crowd mechanism: high disagreement across firms, and individual forecast errors that are widely dispersed but not systematically biased, so that the consensus is far more accurate than the typical respondent? Third, how much does aggregation improve forecast accuracy? The CFO Survey's large and time-varying cross-section allows us to estimate directly how consensus accuracy improves with the number of respondents, and to ask whether those gains are bounded by a component of forecast errors that is common across respondents and therefore cannot be removed by averaging. We do not assume that professional forecasters are more accurate than CFO respondents. They also make systematic errors, so we evaluate the SPF on the same terms as the CFO Survey and test whether its accuracy improves within the calendar year as more data on the current year become available.

We find that the mean subjective expectations from the CFO Survey and the SPF track each other closely, but the two surveys differ sharply in the amount of disagreement across forecasters. Across quarters from 2020Q2 to 2025Q4, CFO disagreement averages about 2.68 percentage points, compared with 0.91 percentage points in the SPF, roughly three times as large. This gap is highly statistically significant and remains present even when we restrict attention to more comparable forecast horizons, such as fourth-quarter surveys forecasting the following calendar year (2.45 versus 0.81 percentage points). Thus, the higher disagreement among CFOs is not simply an artifact of pooling forecasts with different horizons.

Despite this greater disagreement, the aggregated CFO forecasts are competitive with the SPF. The results suggest that the information contained in firm-level probability distributions improve as a result of aggregation (wisdom-of-crowd) and, in some cases, generate forecasts that match or outperform professional forecasters. This pattern is consistent with a wisdom-of-crowd mechanism: individual CFO forecasts are dispersed and uncertain, but aggregation reduces idiosyncratic noise and preserves useful information about aggregate real GDP growth.

Finally, we show that aggregation matters for interpreting SPF performance as well. Individual SPF forecasters can make systematic errors, but these errors are reduced when forecasts are averaged across forecasters. Moreover, SPF accuracy should improve within the calendar year as more information about the current-year base becomes available. However, we fail to find any patterns consistent with learning within calendar years among professional forecasters in the SPF.

2 Related Literature

This paper contributes to a growing literature that uses survey expectations to study how firms, managers, and professional forecasters form beliefs about future economic outcomes. Carroll (2003) provides the foundational account of how non-specialists come to hold macroeconomic views at all: households absorb the forecasts of professionals through news coverage, and because attention to that coverage is intermittent, aggregate expectations adjust sluggishly. Nothing in that mechanism is specific to households, and subsequent work documents the same frictions among firms. Information rigidities are pervasive (Coibion and Gorodnichenko, 2015), firms' expectations are dispersed and systematically related to managerial characteristics and information sets (Coibion et al., 2018), and firm attention to macroeconomic news is heterogeneous enough to affect the transmission of monetary policy (Song and Stern, 2024). The implication for our setting is that the macroeconomic expectations of non-forecasters are neither independent of professional opinion nor a simple replication of it.

A parallel experimental literature tests how non-specialists acquire macroeconomic information, by randomizing what respondents are told. Expectations move in response to central bank communications (Coibion et al., 2022), professional forecasters' recession probabilities (Roth and Wohlfart, 2020), and official inflation statistics (Cavallo et al., 2017). The effects reach decisions as well as beliefs: Coibion et al. (2020) find that randomized inflation information changes Italian managers' pricing, employment, and investment choices. Two features of this evidence bear on our setting. First, respondents weight their own experience heavily relative to official aggregate statistics (Cavallo et al., 2017). Second, attention is not fixed: households and firms become better informed when the macroeconomic environment makes inflation salient (Weber et al., 2025). Business leaders should therefore be expected to hold macroeconomic views that are informed but idiosyncratic. That is the configuration under which aggregation can add value.

A related literature studies business and managerial expectations directly. Altig et al. (2022) show, using the Survey of Business Uncertainty, that firms' subjective distributions over their own sales, employment, and investment predict realized outcomes and forecast errors, and Barrero (2022) finds that managers are overprecise and overextrapolate from recent firm-level shocks. This work establishes firm-level subjective distributions as economically meaningful objects, but it concerns

firms' expectations about their own-firm quantities. We ask whether the same instruments carry information about an aggregate outcome that no single firm experiences.

A separate literature concerns the measurement of probabilistic forecasts and what is lost when beliefs are reduced to point predictions. Manski (2004) makes the general case for eliciting beliefs as subjective probability distributions rather than inferring them from observed choices, and Manski (2018) extends the argument to macroeconomic surveys: a point forecast reports one summary of a respondent's belief while discarding the uncertainty attached to it, and it leaves unstated which feature of that belief, the mean, median, or mode, is being reported. Probability distributions preserve precisely the dispersed information that a consensus point forecast averages away, which is what makes them the right instrument for the question we ask. Turning to professional forecasters specifically, Clements et al. (2023) review the construction of disagreement and uncertainty measures and the calibration of SPF densities, Rich and Tracy (2025) show that forecasters are not interchangeable and that relative performance varies with the target and the difficulty of the environment, and Campbell and Moore (2024) find that SPF forecasts are overprecise, with stated confidence exceeding realized accuracy. This literature treats the professional panel as the object of interest. We instead use it as a benchmark, asking what a very different population of forecasters adds.

Our interpretation of the disagreement gap between the two surveys draws on theories of herding and correlated information. Reputational concerns can generate herd behavior among managers (Scharfstein and Stein, 1990), informational cascades can arise when agents infer information from the actions of others (Bikhchandani et al., 1992), and analysts may issue forecasts closer to prior or consensus expectations than their own information warrants (Trueman, 2015). Clements and Wang (2026) document this directly in the SPF, where the forecasts of others matter for nearly half of respondents' inflation forecasts. Low disagreement among professional forecasters therefore need not indicate superior information, and high disagreement among CFOs need not indicate noise. Nor does it establish private information. Firms may observe the same aggregate conditions as professional forecasters and differ mainly in the precision with which they observe them. That is the noisy-information reading of Coibion and Gorodnichenko (2015) and Coibion et al. (2018) rather than a departure from it. Our aggregation exercise separates the common and idiosyncratic components of CFO forecast errors and finds both to be economically large, indicating that business

leaders do track aggregate conditions while also contributing dispersion that averaging removes. It cannot, on its own, distinguish dispersed private information from heterogeneous processing of common signals, and we do not claim that it can. Either reading requires the distributional data that Manski (2018) advocates, since a consensus point forecast averages away the very dispersion whose informational content is in question.

In related work, Fried et al. (2026) aggregate CFO Survey respondents' expectations of their own firms' price growth into an economy-wide indicator that tracks core CPI inflation and forecasts it more accurately than professional and household surveys. Their exercise is a point-forecast comparison built from firms' expectations about their own outcomes; we compare full predictive distributions over an aggregate that no respondent observes directly, and we treat aggregation itself as the object of study.

We contribute to these literatures in three ways. First, we provide a distributional comparison of business leaders and professional forecasters for the same aggregate target, decomposing the difference between the two surveys into disagreement across respondents and subjective uncertainty within them. Second, we identify the wisdom-of-crowd mechanism directly rather than inferring it from aggregate correlations: by randomizing crowd size within survey waves, we estimate how much of the consensus advantage is attributable to averaging, and we show that these gains are bounded by a component of forecast errors that is common across respondents. Third, we establish what this implies for the use of firm surveys in policy settings. A large firm panel does not purchase accuracy through size alone, but its dispersion is not merely noise, and its consensus is competitive with the SPF at a forecast horizon more directly relevant for real-time policy analysis.

3 Data

This section describes the two main datasets that we use to compare real GDP growth forecasts: the CFO Survey and the SPF. The former relies on firm-level forecasts, while the latter relies on professional forecasters.

3.1 CFO Survey

The CFO Survey is a quarterly survey of US business leaders that elicits the financial outlook for their firms and the US economy.[1] The CFO Survey is conducted as a partnership among Duke University's Fuqua School of Business, the Federal Reserve Bank of Richmond, and the Federal Reserve Bank of Atlanta. It is fielded during the second month of each quarter (February, May, August, and November), and has averaged close to 500 respondents in recent quarters.

The CFO Survey asks respondents for a probability distribution over real GDP growth over the next four quarters (Appendix Figure A1). Respondents allocate probability across 11 bins, ranging from "less than -5.1 percent" to "9 percent or more." The question borrows its format and bins from the SPF but differs in its target: the SPF asks about annual-average growth in the next calendar year relative to the current year (Appendix Figure A2). We compute each respondent's point forecast as the probability-weighted average of the bin midpoints.

Between 2020q2 and 2025q4 (23 quarters), the probabilistic GDP question drew 8,301 individual responses from 1,803 unique respondents, an average of 361 responses per quarter. This question is only asked of respondents that first indicate at least "some" familiarity with GDP, which accounts for much of the difference between the number of total CFO survey submissions and the number of responses to this question (Appendix Figure A3). All responses are required to sum to 100 percent, and respondents cannot proceed unless this condition is met. The bins were last revised in 2024q3, reflecting a similar change to the SPF in 2024q2.

3.2 Survey of Professional Forecasters

The Survey of Professional Forecasters is a quarterly survey of macroeconomic forecasts that began in 1968 and has been managed by the Federal Reserve Bank of Philadelphia since 1990.[2] Results are typically published in February, May, August, and November.

The SPF collects probabilistic expectations for annual-average chain-weighted real GDP growth in the next calendar year relative to the current calendar year. Respondents allocate probability across the same 11 bins described above, and we compute each respondent's probability-weighted

[1] Survey details are available at https://www.richmondfed.org/cfosurvey.

[2] Survey details are available at https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.

average forecast from the bin midpoints.

Between 2020q2 and 2025q4 (23 quarters), the probabilistic GDP question drew 649 individual responses from 54 unique respondents, an average of 28 responses per quarter. Respondents who reported no probabilities are excluded, and all retained responses sum to approximately 100 percent. The bins were last revised in 2024q2.

4 Methodology

This section describes our approach to constructing comparable forecast measures from the CFO Survey and the SPF. Because the surveys elicit full probability distributions rather than point forecasts, we first explain how we derive expectations and uncertainty measures from respondents' probabilistic assessments. We then address a key measurement challenge: the two surveys target slightly different GDP growth concepts — four-quarter growth in the CFO Survey and annual-average-over-annual-average growth in the SPF — so we align forecast horizons.[3] Without this alignment, differences in measured accuracy would partly reflect differences in the target being forecast rather than differences in forecast skill.

4.1 Forecast Construction from Probability Distributions

Both surveys elicit probabilistic forecasts by asking respondents to assign probabilities across discrete bins of real GDP growth outcomes. From these distributions we construct three objects: subjective expectations and subjective uncertainty for each respondent i in quarter t, and disagreement across respondents in each quarter t.

Subjective expectation:

$$\mathbb{E}_{it} = \sum_{j=1}^{11} m_j \cdot p_{itj} \tag{1}$$

Subjective uncertainty (subjective standard deviation):

$$\sigma_{it} = \sqrt{\sum_{j=1}^{11} (m_j - \mathbb{E}_{it})^2 \cdot p_{itj}} \tag{2}$$

[3] In Table 1, we show that in the fourth-quarter version the average difference is about half a percentage point over our sample period: 2.8 percent for four-quarter growth and 3.3 percent for annual-average-over-annual-average growth.

where m_j denotes the midpoint of bin j, and p_{itj} is the probability respondent i in quarter t assigns to bin j.

Disagreement (cross-sectional dispersion of point forecasts):

$$D_t = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\mathbb{E}_{it} - \bar{\mathbb{E}}_t\right)^2}, \qquad \bar{\mathbb{E}}_t = \frac{1}{N_t}\sum_{i=1}^{N_t}\mathbb{E}_{it} \tag{3}$$

where N_t is the number of respondents in survey-quarter t and $\bar{\mathbb{E}}_t$ is that survey's consensus (crowd-mean) forecast. Equations (2) and (3) capture distinct objects: σ_{it} measures uncertainty *within* a respondent's own subjective distribution, whereas D_t measures dispersion *across* respondents' central tendencies.

4.2 Sample Construction and Forecast Horizon Alignment

The two surveys target slightly different GDP growth concepts: the CFO survey asks about four-quarter-ahead growth, while the SPF asks about annual-average-over-annual-average growth. To ensure comparability, we restrict our analysis to forecasts submitted in the fourth quarter of each year. For the SPF, we use respondents' forecasts for the next calendar year (not the current year), which are submitted in Q4 of the preceding year. This timing ensures both surveys are forecasting GDP outcomes approximately one year ahead. The fourth-quarter timing also means the two growth concepts resolve on the same schedule: for forecasts made in the fourth quarter of year t, both four-quarter growth and annual-average-over-annual-average growth are fully realized once one additional year of data, through the fourth quarter of year $t+1$, becomes available.

4.3 Forecast Error Calculation

We evaluate each survey against its corresponding growth target: four-quarter real GDP growth for the CFO Survey and annual-average-over-annual-average real GDP growth for the SPF.[4] Forecast errors are calculated as:

$$e_{it} = y_t^{\text{actual}} - \mathbb{E}_{it} \tag{4}$$

[4]Our findings are robust to using the same target of annual-average-over-annual-average real GDP growth.

where y_t^{actual} is the realized growth rate of the survey's corresponding target concept for the forecast made in quarter t. We use real GDP data as reported by the Bureau of Economic Analysis, without accounting for data revisions.[5]

For each survey and survey quarter t, we construct the consensus forecast as the cross-sectional mean of individual subjective expectations:

$$\mathbb{E}_t^{\text{consensus}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbb{E}_{it} \tag{5}$$

The consensus forecast error is then calculated as $e_t^{\text{consensus}} = \frac{\sum e_{it}}{N_t} = y_t^{\text{actual}} - \mathbb{E}_t^{\text{consensus}}$, where y_t^{actual} is again the realized growth rate of the survey's corresponding growth target.

4.4 Decompositions of Forecast Heterogeneity and Forecast Performance

We perform two decompositions to understand the differences between the forecasts of professionals and business leaders. The first decomposition is ex ante and separates total forecast heterogeneity into two components: uncertainty within individual respondents' subjective probability distributions and disagreement across respondents' subjective expectations. The second decomposition is ex post and separates forecast performance into bias and variance components.

First, using an ANOVA-style variance decomposition, we decompose the total variance of individual real GDP growth estimates from each survey into within-respondent variance and between-respondent variance. In each quarter t, we use each respondent's 11-bin probability distribution to calculate an individual expectation, $\mathbb{E}_{it}$, and an individual variance, $\text{Var}_{it} = \sigma_{it}^2$. The individual expectation is given by Equation (1) where m_j denotes the midpoint of bin j and p_{itj} denotes the probability assigned by respondent i in quarter t to bin j. We then decompose total variance in each quarter as

$$\text{Var}_{\text{total},t} = \underbrace{\mathbb{E}[\text{Var}_{it}]}_{\text{Within: Uncertainty}} + \underbrace{\text{Var}(\mathbb{E}_{it})}_{\text{Between: Disagreement}} . \tag{6}$$

The first term, $\mathbb{E}[\text{Var}_{it}] = \frac{1}{N_t} \sum_{i=1}^{N_t} \text{Var}_{it}$, captures the average subjective variance (within-

[5] As a robustness check, we also evaluate forecasts against the first-release (advance) estimates of real GDP growth; results are qualitatively similar.

respondent variance). The second term, $\text{Var}(\mathbb{E}_{it}) = \frac{1}{N_t}\sum_{i=1}^{N_t}(\mathbb{E}_{it} - \bar{\mathbb{E}}_t)^2 = D_t^2$, captures cross-sectional disagreement across respondents' subjective expectations (between-respondent variance).

This decomposition is purely ex ante because it is based only on respondents' beliefs about future real GDP growth, before realized outcomes are observed. It therefore characterizes the structure of beliefs within each survey: whether forecasters tend to be uncertain but aligned with one another, or relatively confident but in disagreement.

Second, we decompose the mean squared error (MSE) from both surveys to analyze the sources of forecast accuracy. For each period, we use realized real GDP growth and respondents' subjective expectations to calculate individual forecast errors, defined as realized growth minus the forecast. The mean squared error can be decomposed into two components: the squared mean forecast error (bias squared) and the cross-sectional variance of individual forecast errors.

$$\text{MSE}_t = \frac{1}{N_t}\sum_{i=1}^{N_t}(y_t - \mathbb{E}_{it})^2 = \underbrace{\left(y_t - \bar{\mathbb{E}}_t\right)^2}_{\text{Bias}^2} + \underbrace{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\mathbb{E}_{it} - \bar{\mathbb{E}}_t\right)^2}_{\text{Variance}} \tag{7}$$

The bias-squared term captures the consensus error component, while the variance term captures dispersion of individual forecast errors around the survey mean.[6] This decomposition is ex post because it requires realized GDP growth and therefore evaluates forecast performance after outcomes are known.

These two decompositions are linked by an important relationship: the variance component from the ex-post MSE decomposition equals the between-respondent variance from the ex-ante ANOVA decomposition. Specifically, $\text{Var}(\mathbb{E}_{it}) = \frac{1}{N_t}\sum_{i=1}^{N_t}(\mathbb{E}_{it} - \bar{\mathbb{E}}_t)^2$ appears in both Equations 6 and 7.[7] Before outcomes are realized, this quantity measures disagreement across forecasters' subjective expectations. After outcomes are realized, the same quantity measures dispersion of forecast errors around the consensus, since the realized outcome is common to all forecasters in a given survey period. This equivalence implies that disagreement among forecasters at the time they submit their forecasts directly determines part of the ex-post MSE. Even if the consensus forecast exhibits no bias, individual forecast errors will be dispersed to the extent that forecasters initially disagreed

[6] Note that while we express these decompositions in variance terms for mathematical tractability, we report standard deviations elsewhere in the paper to express uncertainty and disagreement in natural units of GDP growth rates.

[7] See Appendix E for the exact derivation.

with one another. This link between disagreement and forecast accuracy should be considered when using accuracy metrics such as MSE, RMSE, or MAE to evaluate individual survey forecasts, as these measures partly reflect dispersion arising from legitimate differences in forecasters' views, particularly when individual forecast errors may largely cancel out in the consensus.

5 Results

In this section, we compare the real GDP growth forecasts of CFO Survey respondents with those of SPF forecasters. Particularly we are interested in measures of forecast accuracy, uncertainty, and how the wisdom of the crowd can lead to useful forecasts from the CFO Survey respondents. Furthermore, we also document systematic errors among professional forecasters in the SPF.

5.1 Forecast Accuracy: CFO vs. SPF

We compare the two surveys along three dimensions, taken in turn: the expectations respondents report, the uncertainty they attach to those expectations, and the errors those expectations produce once outcomes are realized. The first two are ex ante, describing beliefs before the realization is known, and the third is ex post. Table 1 reports pooled summary statistics, and Figure 1 shows the corresponding distributions, with one column of panels for each dimension.[8]

We begin with expectations. The left panels of Figure 1 show that CFO expectations are more dispersed than those of professional forecasters, with the CFO distribution exhibiting wider tails and greater spread around the central tendency. This pattern is reflected in the broader histogram and the wider interquartile range in the box plot. Both surveys nonetheless produce similar mean expectations: 1.89 percent for the CFO Survey and 2.21 for the SPF, with the difference not statistically significant.[9] Individual CFO respondents therefore hold more divergent views about future real GDP growth than the relatively clustered professional forecasters, even though the two groups agree closely on average. Section 5.2 takes up this disagreement gap directly.

[8]Table 1 covers surveys from 2020 to 2024 so that all panels use the same sample. Forecasts made in 2025 target growth in 2026, which is not yet observed. In Figure 1, each measure uses the longest available sample: surveys through 2024 for results that require realized growth, and through 2025 for results on expectations and uncertainty.

[9]The surveys ask about slightly different GDP growth targets, so their mean expectations are not directly comparable. Forecast errors are calculated relative to the realized value of each survey's corresponding target.

Turning to uncertainty, the middle panels of Figure 1 reveal an even more pronounced difference. CFO respondents report higher and more variable uncertainty about real GDP growth, with average subjective uncertainty of 1.63 percentage points compared with 1.49 for SPF forecasters, although the difference is not statistically significant. More importantly, the distribution of CFO uncertainty is highly dispersed: some CFOs report very low uncertainty while others report extremely high uncertainty, as reflected in the long right tail and numerous outliers in the CFO box plot. Professional forecasters, by contrast, exhibit more homogeneous uncertainty, with responses clustering tightly around the median. CFOs thus vary considerably in how predictable they find future real GDP growth, whereas professional forecasters report more similar assessments.

Finally, we turn to realized performance. Forecast accuracy is summarized by three measures: the mean forecast error, which indicates whether forecasts are systematically biased; the mean absolute forecast error, which summarizes the typical size of misses regardless of direction; and the mean squared error, which places greater weight on large mistakes. The right panels of Figure 1 show that CFO forecast errors are far more dispersed than those of professional forecasters, spanning a much wider range with extensive outliers. Table 1 quantifies the gap: mean absolute forecast errors are 2.19 percentage points in the CFO Survey against 1.60 in the SPF, and mean squared errors 10.06 against 4.02, both statistically significant. Mean forecast errors, by contrast, are statistically indistinguishable, differing by 0.18 percentage points, with the CFO consensus marginally closer to the realization. The CFO consensus therefore achieves accuracy comparable to that of professional forecasters despite far greater dispersion in individual errors, a pattern we return to in Sections 5.2 and 5.4.

5.2 Decomposing Forecast Heterogeneity and Comparing Disagreement and Uncertainty between the CFO Survey and the SPF

We begin by applying the ex-ante variance decomposition, Equation (6), described in Section 4.4. Figure 2 shows that professional forecasters and business leaders differ sharply in their ex-ante belief structure. Professional forecasters exhibit lower levels of both disagreement and subjective uncertainty than CFO Survey respondents. Among professional forecasters, subjective uncertainty accounts for most of the total variance, while disagreement across professional forecasters plays a

relatively minor role. For CFO Survey respondents, the opposite pattern emerges: although both components are larger than in the SPF, disagreement across respondents accounts for the larger share of total variance. In this sense, professional forecasters appear relatively aligned but uncertain, whereas CFO Survey respondents appear more divided in their expectations about future real GDP growth. These findings are consistent with the patterns in Figure 1, where CFO forecasts display wider dispersion and thicker tails than SPF forecasts.

The ex-ante variance decomposition establishes which component dominates forecast variance in each survey: disagreement in the CFO Survey and subjective uncertainty in the SPF. However, it does not show the magnitude of the differences between the two surveys, whether these differences persist over time, or whether they remain after accounting for differences in forecast horizons. We examine these questions next by comparing disagreement and subjective uncertainty across the two surveys. This distinction is economically meaningful: disagreement captures heterogeneity in respondents' expectations, whereas subjective uncertainty captures how uncertain the typical respondent is about the outlook.

Disagreement and uncertainty are both dispersion concepts computed from the same probability distributions. Disagreement, D_t in Equation (3), is the cross-sectional standard deviation of individual subjective expectations $\mathbb{E}_{it}$, defined in Equation (1). It captures dispersion *across* forecasters and is the between-respondent component of Equation (6). Said differently, this tells us how much forecasters differ *from each other.* Greater disagreement among CFO respondents may reflect more heterogeneous information sets, differences in attention to aggregate conditions, or the influence of firm-specific information on macroeconomic expectations. An alternative interpretation of lower disagreement among professional forecasters in the SPF is that they are subject to more herding behavior (Clements and Wang, 2026). The standard deviation is the conventional measure in the forecasting literature and uses the full cross-section, although it is sensitive to outliers. While not shown, we verify that the main conclusions are robust to alternative measures of disagreement, including the interquartile range (IQR) and the mean absolute deviation (MAD).

Average subjective uncertainty, $\bar{\sigma}_t = \frac{1}{N_t}\sum_{i=1}^{N_t}\sigma_{it}$, is the cross-sectional mean of the subjective standard deviations defined in Equation (2). It captures dispersion *within* forecasters and tells us how unsure the *typical forecaster* is. Its counterpart in variance terms, the average subjective variance $\frac{1}{N_t}\sum_{i=1}^{N_t}\sigma_{it}^2$, is the within-respondent component of Equation (6). Greater uncertainty

among CFO respondents may suggest a more noisy environment with less informative data signals for this group of respondents. It would imply that CFOs as a group interpret the environment as more uninformative about GDP growth compared to professional forecasters. Conversely, greater uncertainty may reflect higher aggregate volatility about the macroeconomic environment. Rather than reflecting more common noise, it may reflect a less predictable environment (as may arise around recessions, policy regime shifts, pandemics, etc.).

An important feature of the SPF comparison is that the following-year forecast is a fixed-target object whose horizon shrinks over the course of the survey year. In the first quarter, SPF respondents are forecasting annual-average over annual-average real GDP growth roughly 21 months ahead, while in the fourth quarter the object that professional forecasters are still forecasting is only 12 months ahead, similar to the CFO Survey. We therefore focus on the fourth-quarter SPF forecasts.[10]

Although mean expectations in the two surveys track each other closely, disagreement does not. In fourth-quarter surveys, CFO disagreement averages 2.45 percentage points against 0.81 for the SPF, roughly three times as large. The gap is persistent: CFO disagreement exceeds SPF disagreement in every survey quarter from 2020Q2 to 2025Q4 (Appendix Figure A7).

The right panel of Figure 3 shows fourth-quarter disagreement for each survey from 2020 through 2025.[11] Business leaders hold substantially more dispersed beliefs about GDP growth than professional forecasters, consistent with greater heterogeneity in their underlying information sets. At the same time, disagreement in both surveys rises and falls with the same broad uncertainty cycle.

In the left panel of Figure 3, we compare subjective uncertainty between the CFO Survey and SPF respondents. Average subjective uncertainty, $\bar{\sigma}_t$, is 1.50 percentage points for the SPF against 1.64 for the CFO survey across fourth-quarter surveys (1.56 versus 1.69, respectively, using all surveys). Both series decline steadily from their 2020–2021 pandemic peaks and track one another closely over the entire sample. Average subjective uncertainty is only about 9 percent higher in the CFO Survey than in the SPF (8 percent across all survey quarters), yet CFO disagreement is roughly three times as large as SPF disagreement. The greater dispersion of CFO expectations is thus almost entirely a between-respondent phenomenon rather than a within-respondent one. Business leaders

[10] Results using all survey quarters are similar and reported in Appendix Figure A7.

[11] These values differ from Table 1, which pools respondents across survey quarters before computing disagreement.

do not report substantially greater subjective uncertainty about the outlook; they hold substantially more heterogeneous central expectations than professional forecasters. This divergence between disagreement and uncertainty also bears on the calibration of subjective uncertainty, which we examine in Appendix C.

5.3 Decomposing Consensus Accuracy and Comparing Bias and Dispersion of Forecast Errors

Turning to the ex-post MSE decomposition, Equation (7), Figure 4 shows that CFO respondents have a higher variance component but a bias-squared component similar to that of professional forecasters. This implies that professional forecasters are more precise at the individual level, as their forecast errors are less dispersed across respondents. However, the similar bias component means that the average CFO forecast performs similarly to the average SPF forecast, leading to comparable consensus-level accuracy. These results are consistent with the right panels of Figure 1, where CFO forecast errors are centered similarly but far more dispersed, and with Panel B of Table 1, where the CFO Survey has a slightly smaller mean error but a substantially larger error standard deviation and MSE. Consensus forecast errors are statistically indistinguishable across the two surveys, with a difference of 0.18 percentage points, and are also statistically indistinguishable in four of the five quarters individually. By contrast, mean squared errors are higher for CFO respondents, with a difference of 6.04 (statistically significant at the 5 percent level). Under the decomposition in Equation (7), this implies that the higher MSE among CFO respondents is driven by greater forecast-error variance rather than consensus bias. As we established in 4.4, the error variance is the same component as the between-firm variance from the ANOVA decomposition and is determined at the time forecasts are made. This pattern is consistent with a wisdom-of-crowd interpretation: although individual CFO forecasts are noisier, idiosyncratic errors appear to offset one another in aggregation. This distinction is important because survey forecasts are typically used in aggregate form, making consensus accuracy especially relevant for evaluating their forecasting performance.

Figure 5 provides a more direct view of the two components of Equation (7), plotting the mean and standard deviation of forecast errors in percentage points. The left panel shows that

mean forecast errors track closely across the two surveys throughout the sample, with similar patterns of over- and under-prediction across quarters. The largest divergence occurs in 2020Q4, when CFO respondents under-predicted GDP growth by a larger margin than SPF forecasters, indicating that their forecasts anticipated a weaker recovery. In contrast, the right panel shows that the standard deviation of CFO forecast errors is consistently two to three times larger than that of the SPF. CFO forecast-error standard deviations range from roughly 1.5 to 4 percentage points, compared with about 0.5 to just over 1 percentage point for professional forecasters. Thus, the higher CFO MSE reflects greater dispersion of individual forecast errors rather than larger consensus bias. The persistence of this dispersion gap across quarters reinforces the view that aggregation (wisdom-of-crowd) substantially reduces the idiosyncratic noise in CFO forecasts.

5.4 Wisdom of the Crowd in the CFO Survey

A natural question is whether the CFO consensus forecast benefits from the survey's large cross-section: if individual errors are at least partly idiosyncratic, averaging over more respondents should shrink the error of the consensus. The law of large numbers bounds how large this effect can be. If forecast errors were independent across respondents, the standard deviation of the consensus error would decline at rate $N^{-1/2}$, so the elasticity of the crowd's absolute error with respect to crowd size is bounded below by -0.5; any common component of errors makes the true elasticity strictly smaller in magnitude. In the raw time series, both mean forecast errors and mean absolute forecast errors in the CFO Survey improve over the sample period, coinciding with growth in survey participation (from 263 forecasters at the trough to 483 at the peak). One concern, however, is that this pattern reflects changes in the aggregate economy rather than aggregation itself. Uncertainty spiked at the onset of the pandemic in 2020q2, when participation was at its trough, and was comparatively subdued from 2021 through 2024, over which participation grew.[12]

Formally, let e_{it} denote the forecast error of respondent i in quarter t (realized annual real GDP growth minus the subjective expectation), and let N_t be the number of respondents in wave t. Because the realization is common to all respondents within a quarter, the mean error $\mathrm{MFE}_t = N_t^{-1} \sum_i e_{it}$ equals the forecast error of the consensus (crowd-mean) forecast, while $\mathrm{MAFE}_t = N_t^{-1} \sum_i |e_{it}|$ measures the accuracy of the average individual. Columns 1–2 and 4–5 of Table 4

[12] Most of the increase in participation came after 2022.

estimate the time-series regression

$$y_t \;=\; \alpha \;+\; \beta \log N_t \;+\; \gamma \log(\mathrm{EPU}_t) \;+\; \varepsilon_t, \qquad y_t \in \big\{\log|\mathrm{MFE}_t|,\ \log \mathrm{MAFE}_t\big\}, \tag{8}$$

where EPU_t is the economic policy uncertainty index of Baker et al. (2016), included to control for the forecasting environment; the EPU control is omitted in the odd columns. The first two columns, where the outcome variable is $\log|\mathrm{MFE}_t|$, deliver elasticities of -3.8 and -3.0 — six to eight times the theoretical bound. Taken literally, these estimates would imply that doubling the panel cuts the consensus error by more than 85 percent, which no aggregation mechanism can produce. The last two columns, where the outcome variable is $\log \mathrm{MAFE}_t$, confirm that the raw time-series relationship is confounded, through what amounts to a placebo test. Aggregation operates only on $|\mathrm{MFE}_t|$, the error of the average forecast; it cannot affect MAFE_t, the average of the individual errors, since that quantity is a within-wave mean of $|e_{it}|$ and does not depend on how many respondents are averaged together. Under a pure aggregation story the elasticity of MAFE with respect to N should therefore be zero: crowd size changes how well the crowd forecasts, not how well its typical member does. Instead, MAFE falls nearly as steeply as the consensus error, with elasticities of -2.1 and -1.6 against -3.8 and -3.0. Most of the apparent gain from a larger crowd is thus a property of the quarters themselves rather than of the crowd: waves with many respondents were simply easier quarters to forecast, so panel size and forecast difficulty move together in the time series.

To isolate the aggregation effect, column 3 turns to within-quarter randomization. For each survey wave we draw 200 random subsets of $k \in \{2, 5, 10, 15, 20, 30, 40\}$ respondents without replacement, compute the absolute error of each subset's consensus forecast, and average across draws. Denoting by $\overline{|\mathrm{MFE}|}_{t,k}$ the resulting average absolute consensus error in the quarter-by-k panel, we estimate

$$\log \overline{|\mathrm{MFE}|}_{t,k} \;=\; \alpha_t \;+\; \beta \log k \;+\; \varepsilon_{t,k}, \tag{9}$$

where the quarter fixed effects α_t ensure that small and large crowds are compared forecasting the same quarter with the same information, so the estimate is purged of anything that varies at the wave level; standard errors are clustered by quarter. The resulting elasticity is $\hat{\beta} = -0.17$ (s.e. 0.041): a genuine wisdom-of-crowd effect operates in the CFO Survey, with a doubling of the crowd reducing

the consensus error by about 11 percent, and a move from a two-person to a forty-person crowd reducing it by roughly 40 percent.

The magnitude, however, is well below the independence benchmark: a Wald test rejects an elasticity of $-1/2$ ($p < 0.001$), just as it rejects the implausibly steep time-series estimates from the opposite direction. The gap is informative about the structure of CFO errors. Under independence, doubling the crowd should eliminate 29 percent of the consensus error; the estimated 11 percent implies that most of the consensus error is a common component, reflecting shared macroeconomic shocks and shared narratives, that averaging cannot remove. This is itself informative about attention: a common component of this size means that CFO respondents are responding to the same aggregate developments as one another, not forecasting from firm-specific information alone. Aggregation gains are therefore real but quickly exhausted. Interpreted through a two-component error model, the estimated elasticity implies that idiosyncratic variance is roughly six and half times common variance, so the consensus error sits within about 10 percent of its common-component floor once a wave reaches 30 respondents, within 5 percent by roughly 60 respondents, and within 1 percent at the CFO Survey's typical wave size of about 360.[13] Further growth in the panel yields little additional accuracy: moving from an SPF-sized crowd of about 28 respondents to a CFO-sized crowd of about 360 reduces the consensus error by less than 10 percent. This has two implications for our setting. First, the CFO Survey's large cross-section cannot, by crowd size alone, close an accuracy gap with the much smaller SPF panel, since consensus accuracy is ultimately bounded by the common component of errors rather than by the number of respondents. Second, the contrast between the naive elasticity of -3.8 and the experimental estimate of -0.17 is a caution against interpreting time-series correlations between survey participation and forecast accuracy causally.

5.5 Within-Year Learning in the SPF

One concern with using CFO Survey expectations is that business leaders may face greater forecasting challenges than professional forecasters. CFOs may have less specialized macroeconomic

[13] Write $e_{it} = c_t + u_{it}$, where c_t is an error component common to all respondents and u_{it} is idiosyncratic noise. The consensus error then has variance $\sigma_c^2 + \sigma_u^2/N$, so its standard deviation relative to the floor σ_c is $\sqrt{1 + r/N}$ with $r = \sigma_u^2/\sigma_c^2$, and the elasticity with respect to crowd size equals $-\frac{1}{2}$ times the idiosyncratic share of consensus-error variance. Evaluating the estimated elasticity of -0.17 at the geometric mean of the randomization grid ($k \approx 12$) yields $r \approx 6.4$, which delivers the convergence schedule in the text.

expertise, update less frequently, or place greater weight on firm-specific information. Each of these amounts to the same claim: that CFOs allocate less attention to aggregate macroeconomic information than professional forecasters do, whether by constraint or by choice. Before attributing the dispersion in CFO forecasts to a difference in attention, we test whether professional forecasters themselves display the attentiveness CFOs are presumed to lack. The SPF's fixed-target structure makes this directly observable: efficient forecasters should become more accurate within the year as data on the current year arrive.[14]

Forecasts made in Q1 through Q4 of year Y_t all refer to the same object: annual-average real GDP growth in year Y_{t+1} relative to year Y_t. As the year progresses, forecasters observe more information about the current-year path of GDP and the annual-average base, so more information about the target is available by construction. If professional forecasters use it efficiently, absolute forecast errors should decline from Q1 to Q4.

The results provide little evidence of systematic within-year improvement. Table 2 reports a linear-trend regression of absolute forecast errors on survey quarter, coded from 1 to 4, with and without year fixed effects and standard errors clustered by year. The quarter coefficient is positive in both specifications and statistically insignificant, indicating no clear decline in forecast errors over the year. If anything, the point estimates go in the opposite direction, suggesting that SPF forecasters do not reduce their absolute forecast errors as the calendar year progresses. Table 3 provides complementary sequential quarter-to-quarter contrasts. The Q1 to Q2 transition shows a deterioration in forecast accuracy of 0.45 percentage points in the specification with year fixed effects, significant at the 10 percent level, while the Q2 to Q3 and Q3 to Q4 transitions are small and statistically insignificant. These results are consistent and suggest limited within-year learning among SPF forecasters.

Taken together, these results show that even professional forecasters do not consistently improve their annual-average-over-annual-average real GDP forecasts as additional within-year information arrives. This finding tempers concerns that greater dispersion in CFO forecasts reflects a problem unique to business leaders. More broadly, it reinforces the importance of distinguishing individual-level forecast errors from consensus-level performance: even when individual forecasts are noisy or

[14]The test is available only for the SPF: the CFO Survey's rolling four-quarter-ahead horizon means successive waves forecast different objects, leaving no fixed target against which learning could be measured.

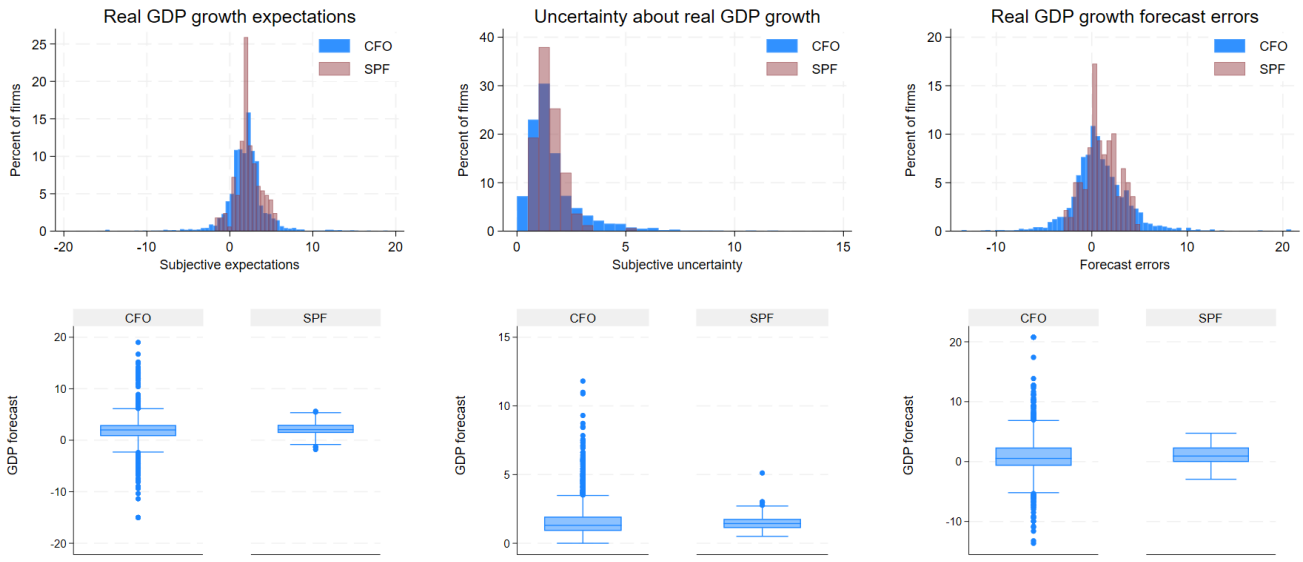
imperfectly updated, aggregation can offset idiosyncratic errors and produce a more informative consensus forecast.

6 Conclusion

These findings suggest that while CFO forecasters do not outperform professional forecasters in terms of consensus accuracy, there is also no evidence that they underperform. The MSE decomposition reveals that both surveys achieve similar consensus-level accuracy, with the CFO survey's higher mean squared error stemming entirely from greater variance in individual forecasts rather than consensus over/under-prediction. The ANOVA decomposition reveals that CFO forecasts are characterized by high disagreement (between-respondent variance dominates), meaning CFOs hold divergent views about future GDP growth. When these diverse forecasts are evaluated against actual outcomes, this disagreement translates into dispersed forecast errors—reflected in the high variance component of the MSE decomposition. However, as demonstrated by the similar consensus accuracy, this error dispersion is largely idiosyncratic (averaging to near-zero bias), allowing the wisdom-of-crowd mechanism to work effectively: diverse but unbiased individual forecasts aggregate to an accurate consensus. Our within-quarter randomization exercise provides direct evidence of this wisdom-of-crowd effect: comparing randomly drawn crowds of different sizes forecasting the same quarter with the same information, a larger crowd produce systematically more accurate consensus forecasts—moving from a two-person to a forty-person crowd reduces the consensus error by roughly 40 percent—though gains flatten quickly. Given this comparable performance at the consensus level, the CFO survey presents a valuable complement to the SPF, particularly because the two surveys target different forecast horizons. While the SPF focuses on annual-average-over-annual-average GDP growth, the CFO survey elicits forecasts of four-quarter-ahead growth—a measure that may be more relevant for real-time policy analysis and business decision-making, as it captures growth dynamics without the temporal smoothing inherent in annual averages. The availability of both surveys thus provides policymakers and researchers with a richer information set spanning different forecaster types, information sets, and time horizons.

Tables and Figures

Figure 1: Characteristics of Survey Forecasts



Notes: This figure includes histograms and box plots of subjective expectations, uncertainties and forecast errors of future real GDP growth. For comparability, the sample includes only estimates provided in the fourth quarter of each year. The sample period for expectations and uncertainty plots spans from 2020:Q4 to 2025:Q4, while the sample period for the forecast error plots spans from 2020:Q4 to 2024:Q4.

Table 1: Summary Statistics

	CFO Survey		SPF (PRGDP)	
	Mean	SD	Mean	SD
Panel A: Realized outcomes and point forecasts				
Realized GDP growth[†] (p)	2.81	1.41	3.34	1.48
Point forecast (p)	1.96	2.62	2.31	1.57
Panel B: Forecast accuracy				
Forecast error (pp)	0.85	3.06	1.03	1.73
Abs. forecast error (pp)	2.19	2.29	1.60	1.22
RMSE (pp)	3.17	—	2.01	—
MSE (pp^2)	10.06	—	4.02	—
Panel C: Uncertainty				
Subjective uncertainty (pp)	1.76	1.39	1.55	0.63
Forecast variance (pp^2)	5.01	—	2.80	—
Observations	1,716		139	
Quarters (Q4 only)	5		5	

Notes: Individual-level observations, Q4 surveys only, 2020–2024; forecasts target real GDP growth one year ahead, so realized outcomes cover 2021–2025 (only affects Realized GDP growth). All statistics are computed by pooling individual responses across all five survey quarters, rather than computing each statistic within a quarter and then averaging. Realized GDP growth is constant within a survey quarter, so its reported standard deviation reflects variation across quarters rather than across respondents. Point forecasts and subjective uncertainty are computed from probability-bin responses using expected-midpoint (EM) bin midpoints with a bin break in 2024q2. Forecast error is realized GDP growth minus the individual point forecast, and RMSE is the square root of the mean squared forecast error. Subjective uncertainty is each respondent's implied standard deviation from their probability distribution over bins. Within-respondent variance (forecast variance) is the average across respondents of the squared subjective uncertainty, and is the within-respondent component of the ex-ante variance decomposition.

[†] For the SPF, realized GDP is the annual-average-over-annual-average (AAoAA) growth rate, matching the PRGDP elicitation. For the CFO Survey, it is the Q4-over-Q4 growth rate. For a survey in year t, both are measured over year $t + 1$.

Table 2: Within-Year Learning Test: SPF PRGDP Forecasters

	Abs. Forecast Error	Abs. Forecast Error
Quarter (1–4)	0.173	0.076
	(0.133)	(0.152)
Year FE	No	Yes
N	538	538

Notes: The dependent variable is the absolute forecast error for annual real GDP growth, measured in percentage points. This is using SPF individual probability data for Real GDP growth from 2020q2 to 2024q4. The variable Quarter is coded from 1 to 4 within each calendar year, so a negative coefficient indicates within-year improvement in forecast accuracy. Column (2) includes year fixed effects. Standard errors, clustered by year, are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Within-Year Quarter Transitions: SPF PRGDP Forecasters

Transition	Abs. Forecast Error (1)	Abs. Forecast Error (2)
Q1 $\rightarrow$ Q2	0.842	0.454*
	(0.452)	(0.194)
Q2 $\rightarrow$ Q3	-0.002	0.045
	(0.099)	(0.096)
Q3 $\rightarrow$ Q4	-0.197	-0.210
	(0.280)	(0.290)
Year FE	No	Yes
N	538	538

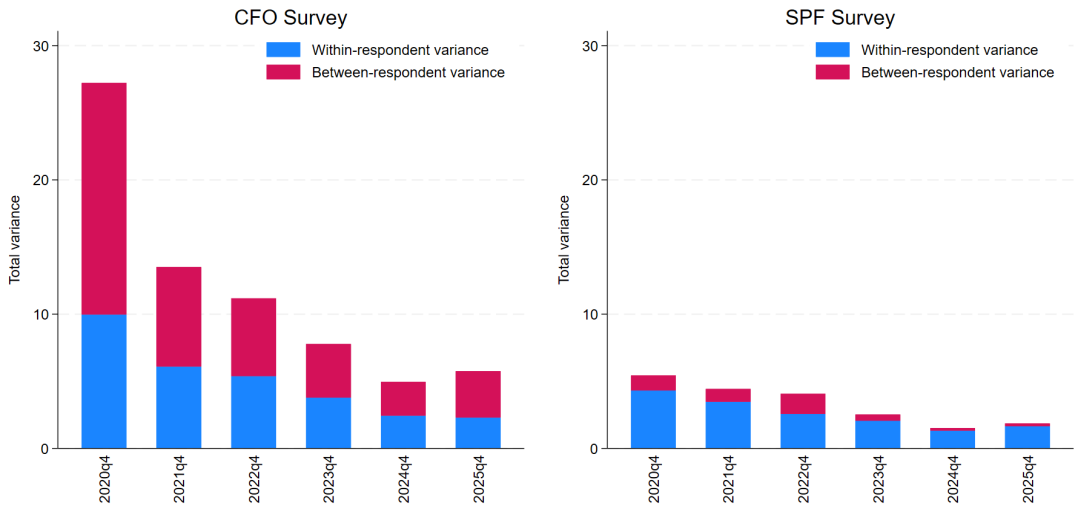
Notes: The dependent variable is the absolute forecast error for annual-average real GDP growth, measured in percentage points, using SPF individual probability data from 2020q2 to 2024q4. Each row reports the estimated change in mean absolute forecast error between adjacent survey quarters within the same calendar year (e.g., Q1 $\rightarrow$ Q2 is the Q2 mean minus the Q1 mean), so a positive coefficient indicates that forecast accuracy deteriorates from one quarter to the next. Column (1) reports unconditional contrasts; column (2) adds year fixed effects. Standard errors, clustered by year, are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Wisdom of Crowds: Forecast Accuracy and Number of CFO Respondents

	Crowd accuracy			Individual accuracy	
	$log(\|MFE\|)$	$log(\|MFE\|)$	$log(\|MFE\|)$	$log(MAFE)$	$log(MAFE)$
log(N)	-3.803***	-2.978***	-0.174***	-2.079***	-1.563***
	(1.028)	(0.781)	(0.041)	(0.388)	(0.338)
log(EPU index)		1.366**			0.853**
		(0.534)			(0.327)
Quarter FE	No	No	Yes	No	No
Observations	20	20	140	20	20

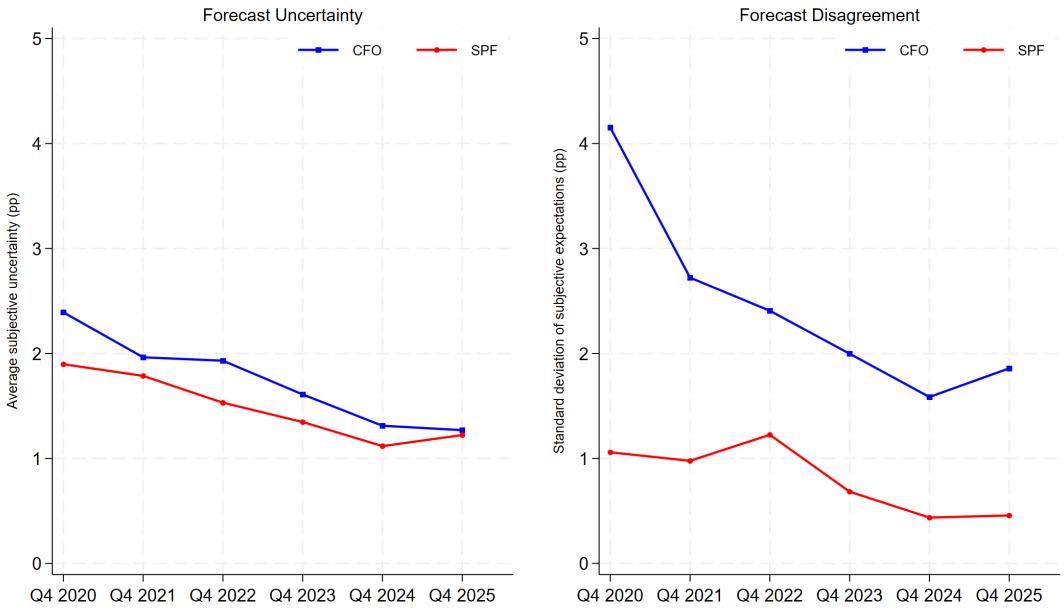
Notes: The dependent variable is the log of the absolute value of the mean forecast error in a given quarter (columns 1–3) and the log of the mean absolute forecast error (columns 4–5), using CFO Survey individual probability data for real GDP growth from 2020q2 to 2025q1. In columns 1–2 and 4–5, the unit of observation is a survey quarter and N is the number of forecasters in the corresponding survey wave; robust standard errors are reported in parentheses. Column 3 is estimated on within-quarter random subsamples: for each quarter and each subsample size $k \in \{2, 5, 10, 15, 20, 30, 40\}$, we draw 200 random subsets of k respondents without replacement and average the absolute error of each subset's consensus forecast; $\log(N)$ there refers to the subsample size k, the regression includes quarter fixed effects, and standard errors are clustered by quarter. The EPU index is the Economic Policy Uncertainty index of Baker et al. (2016). * p<0.10, ** p<0.05, *** p<0.01.

Figure 2: Total Variance Decomposition (ANOVA) of Individual Real GDP Growth Estimates



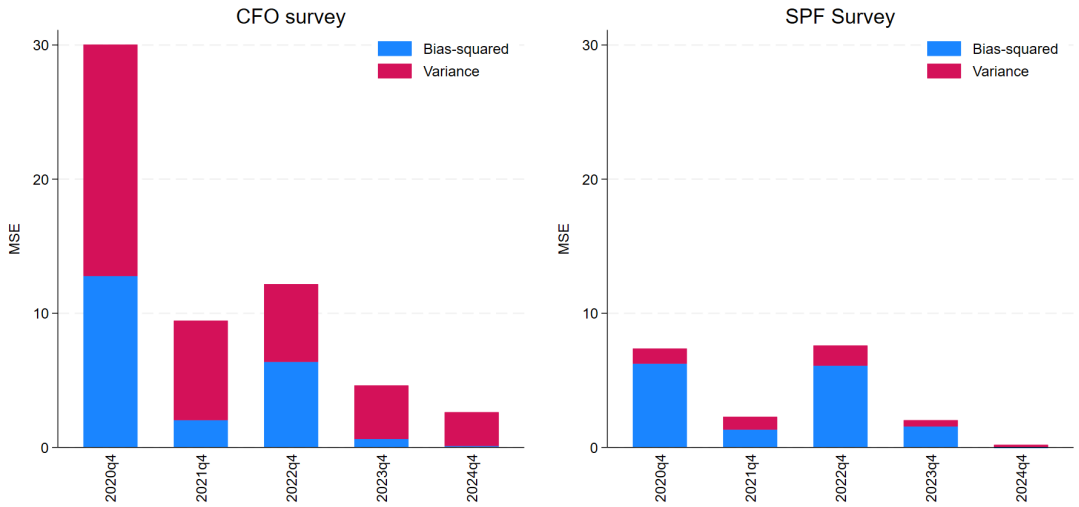
Notes: This figure represents the variance decomposition (ANOVA) of individual probability-weighted real GDP growth estimates. Total variance is decomposed into within-respondent variance (average subjective variance) and between-respondent variance (disagreement across respondents' point forecasts). For comparability, the sample includes only estimates provided in the fourth quarter of each year.

Figure 3: Forecaster Uncertainty and Disagreement (Q4 Surveys Only)



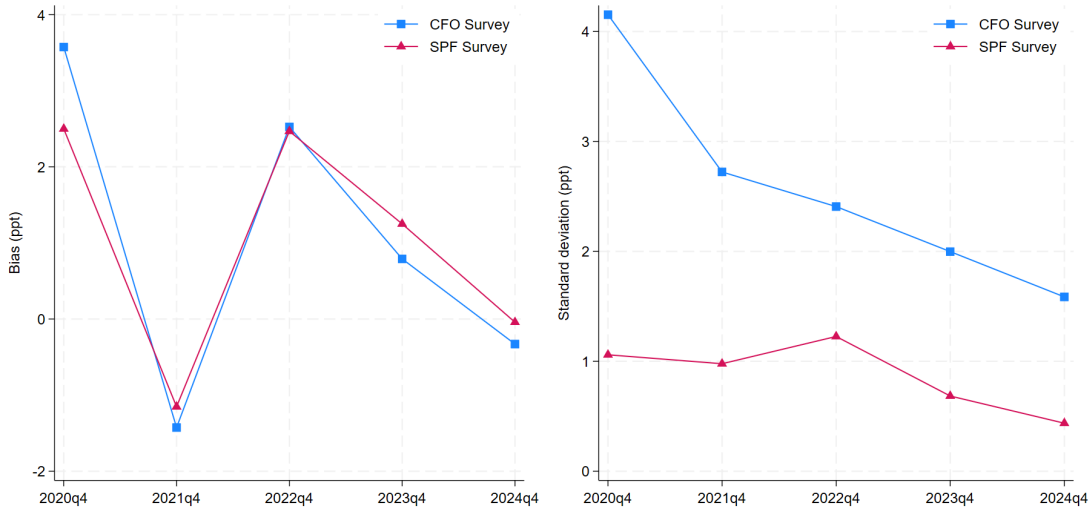
Notes: The left panel plots average subjective uncertainty, measured as the mean across respondents of the standard deviation of each respondent's subjective probability distribution for real GDP growth. The right panel plots disagreement, measured as the cross-sectional standard deviation of respondents' point forecasts. Both measures are expressed in percentage points. The figure uses fourth-quarter surveys only, when the SPF following-year forecast has an approximately 12-month horizon and therefore most closely matches the CFO Survey's year-ahead forecast horizon. Respondents who report no probabilities are excluded from the uncertainty measure, and all retained probability distributions sum to approximately 100 percent.

Figure 4: MSE Decomposition of Forecast Errors



Notes: This figure represents the decomposition of mean squared forecast errors (MSE) into two components: squared bias (the squared mean forecast error, capturing the average directional error in that period) and variance (the dispersion of individual forecast errors around their mean, capturing inconsistency across forecasters). For comparability, the sample includes only forecasts provided in the fourth quarter of each year.

Figure 5: Comparison of Bias and Dispersion of Forecast Errors



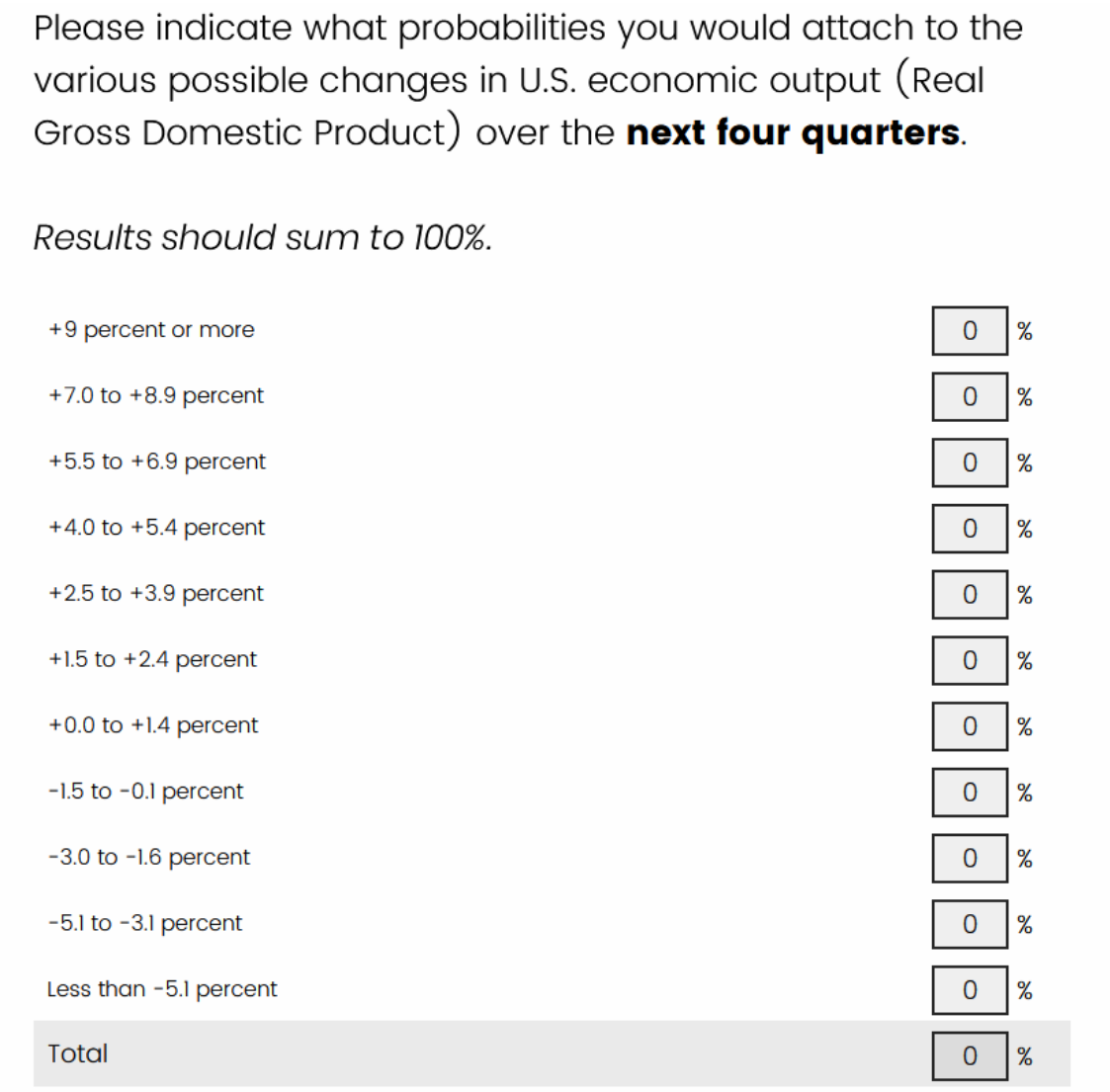
Notes: The left panel of this figure shows the mean forecast errors (bias) by quarter for each survey. The right panel shows the standard deviation of individual forecast errors within each quarter. For comparability, the sample includes only forecasts provided in the fourth quarter of each year.

Appendix

A Survey Questions

We show the survey questions behind our forecast measures, described in Sections 3.1 and 3.2. Figure A1 shows the CFO Survey question on real GDP growth. It asks respondents to assign probabilities to 11 bins for growth over the next four quarters, and the probabilities must sum to 100 percent. Figure A2 shows the corresponding SPF question. It uses the same 11 bins but asks about annual-average over annual-average growth in chain-weighted real GDP, for the current year and each of the next three years. This difference in target is why we align forecast horizons in Section 4. Figure A3 shows the CFO Survey question on familiarity with GDP growth statistics. Only respondents who report at least some familiarity see the probability question.

Figure A1: CFO Survey Real GDP Growth Expectations Question



Notes: The question as shown to respondents beginning in 2024Q3, after the most recent bin revision.
Source: The CFO Survey.

Figure A2: SPF Real GDP Growth Expectations Question

Section 4A. Probabilities of Changes in Real GDP (Y/Y)

Please indicate what probabilities you would attach to the various possible percentage changes (annual-average over annual-average) in chain-weighted real GDP. The probabilities of these alternative forecasts should add up to 100. For the last two years, please provide answers only if you normally forecast out that far.

	Probability of indicated percent change in real (chain-weighted) GDP			
	2023-2024	2024-2025	2025-2026	2026-2027
9.0 percent or more				
7.0 to 8.9 percent				
5.5 to 6.9 percent				
4.0 to 5.4 percent				
2.5 to 3.9 percent				
1.5 to 2.4 percent				
0.0 to 1.4 percent				
-1.5 to -0.1 percent				
-3.0 to -1.6 percent				
-5.1 to -3.1 percent				
Decline more than 5.1 percent				
TOTAL	0	0	0	0

Notes: Section 4A of the 2024Q2 SPF survey form. *Source:* Federal Reserve Bank of Philadelphia, Survey of Professional Forecasters.

Figure A3: CFO Survey GDP Familiarity Question

Notes: Respondents answer this question before the real GDP growth question in Figure A1. *Source:* The CFO Survey.

B Analysis of Top Forecasters in the CFO Survey

A natural question is whether the CFO Survey's consensus accuracy is driven by a small set of consistently superior forecasters—"superforecasters" whose performance systematically exceeds that of the broader panel. If such forecasters exist and can be identified, forecast accuracy could be improved by emphasizing their predictions or understanding the distinctive methods that underlie their success. To assess this possibility, we examine two complementary questions: first, whether forecasting accuracy is persistent across individuals (do the most accurate forecasters in one quarter remain accurate in the next?), and second, whether forecasters identified as top performers contribute disproportionately to the survey's overall accuracy.

We begin by tracking the stability of top forecasters over consecutive quarters. For each survey wave, we identify the forecasters with the lowest absolute forecast errors—specifically, those whose errors place them at or below the 30th-ranked forecaster, a threshold that includes ties and typically yields 30–40 forecasters per quarter. Among this group, we calculate the share who were also top performers in the immediately preceding quarter, conditioning on participation in both waves. Figure A4 plots this persistence rate over time. The share averages approximately 15–20 percent and exhibits substantial variation across quarters, ranging from below 10 percent in some periods to above 30 percent in others. This low persistence rate suggests that forecast accuracy in the CFO Survey is not consistently concentrated among the same individuals; rather, the set of top performers rotates considerably from quarter to quarter, consistent with accuracy being driven more by idiosyncratic luck or quarter-specific information advantages than by stable forecasting skill.

To assess whether top performers contribute meaningfully to aggregate accuracy when they can be identified, we adopt a forecaster-ranking approach based on cumulative performance. For each quarter beginning in 2021Q2, we identify forecasters in the top 20 percent (top two deciles) of cumulative average absolute forecast errors, requiring at least four prior survey submissions to ensure rankings reflect a meaningful track record. We then evaluate whether forecasters flagged as top performers in quarter $t-1$ exhibit lower forecast errors in quarter t relative to the remainder of the panel. Figure A5 presents two complementary views of this comparison. The left panel plots the time series of mean absolute forecast errors for top-tier forecasters (identified in the previous quarter) alongside the full panel average. The two series track closely throughout the evaluation period, with

top forecasters' errors hovering near—and occasionally exceeding—the panel average. The right panel displays p-values from two-sample t-tests comparing top-tier and non-top-tier accuracy in each quarter; the majority of p-values lie well above conventional significance thresholds (insignificant at 10 percent level), and several exceed 0.30. Across the 15 quarters examined, top forecasters significantly outperform the remaining panel in only a handful of cases. These results indicate that ex-ante identification of superior forecasters, even when based on sustained prior performance, does not reliably translate into superior subsequent accuracy.

Taken together, these findings support the interpretation that the CFO Survey's consensus accuracy arises through aggregation rather than through the influence of a few exceptional individuals. The low quarter-to-quarter persistence in top-forecaster identity and the absence of significant performance gaps between ex-ante top performers and the broader panel are consistent with our earlier variance-decomposition results: individual CFO forecast errors are highly dispersed (reflecting substantial idiosyncratic variance), but the consensus forecast achieves accuracy comparable to professional forecasters because idiosyncratic errors largely cancel in the aggregate. This aligns with the classic wisdom-of-crowds mechanism, in which a diverse set of individually noisy forecasts aggregates to a more accurate consensus, provided individual errors are not dominated by a common systematic component. The evidence suggests that the CFO Survey's value lies in its breadth of participation rather than in the predictive power of a select few, reinforcing the importance of maintaining a large and heterogeneous panel rather than attempting to identify and weight a small subset of forecasters.

C Overprecision in the SPF and CFO Surveys

C.1 Measure

A forecaster is *overprecise* when the subjective distribution they report is too narrow relative to the errors their forecasts actually make (Moore and Healy, 2008). Both surveys elicit a probability distribution over year-ahead annual real GDP growth. From each respondent's distribution we take two objects: the point forecast, computed as the probability-weighted average of bin midpoints, and the subjective standard deviation—the spread of that same distribution, which measures the

uncertainty the respondent admits. We then score these forecasts against realized outcomes: a survey taken in year t targets the following calendar year, so we compare each survey's forecast RMSE against actual annual growth to the average subjective standard deviation it reported. Their ratio is the overprecision measure,

$$\text{overprecision} \;=\; \frac{\text{RMSE of point forecasts}}{\text{average subjective SD}},$$

which equals one when densities are correctly sized and exceeds one when realized errors outrun the uncertainty forecasters acknowledged. We use fourth-quarter surveys, where the year-ahead horizon is fixed and the two surveys' targets align most closely, and score 2020–2024 surveys against 2021–2025 outcomes.

C.2 Findings

Both groups are overprecise, and CFOs substantially more so (Figure A6). Pooling across years, realized errors exceed admitted uncertainty by about a third for professional forecasters and about four-fifths for CFOs; equivalently, subjective standard deviations are too small by roughly a quarter and a half, respectively.

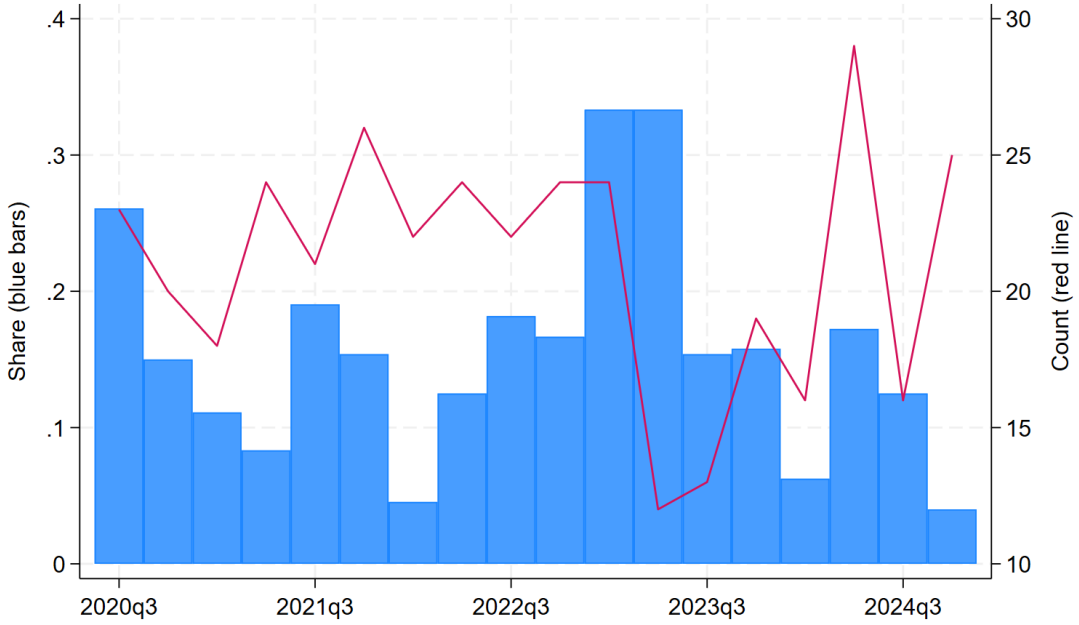
The year-by-year pattern shows this is pervasive rather than the artifact of a single episode. The CFO ratio exceeds one in every target year, peaking in 2021 when the pandemic recovery (6.2 percent growth) fell far above any forecast. Professional forecasters are more variable: overprecise in 2021, 2023, and 2024, but adequately wide—even a touch cautious—in the calmer years 2022 and 2025. The CFO ratio exceeds the SPF ratio in four of the five years. Because each survey contributes only one fourth-quarter observation per year (with roughly 25–30 SPF respondents), the individual SPF bars are noisy and are best read as indicative; the pooled contrast is firm.

D Extension Beyond Q4 Surveys

D.1 Disagreement and Subjective Uncertainty

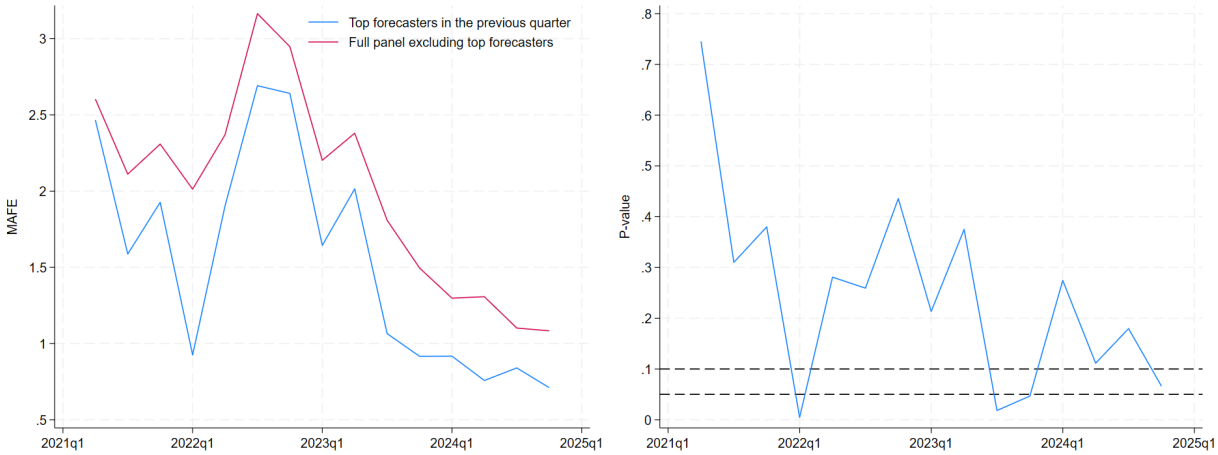
In the main text, we compare forecast disagreement and subjective uncertainty using only fourth-quarter surveys, as shown in Figure 3, to better align forecast horizons between the CFO Survey and the SPF. Here, we extend the analysis to all survey quarters. Figure A7 shows that our main findings are unchanged: both disagreement and subjective uncertainty are higher in the CFO Survey than in the SPF, but the difference between the two surveys is substantially larger for disagreement than for subjective uncertainty. When we consider all quarters from 2020Q2 to 2025Q4, CFO disagreement averages 2.69 percentage points, compared with 0.91 percentage points for the SPF—roughly three times as large—and the gap is highly significant, with CFO disagreement persistently higher.

Figure A4: Persistence of Top Forecasters in Consecutive Surveys



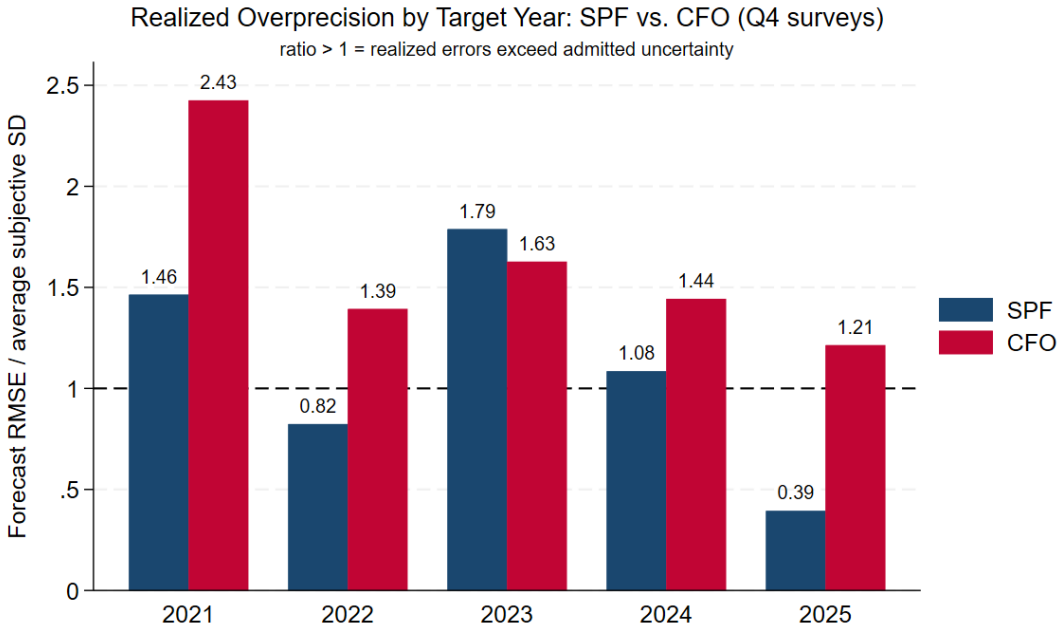
Notes: This figure shows the share of top forecasters in quarter t who were also top forecasters in quarter t-1. Top forecasters are defined as those with absolute forecast errors at or below the 30th-ranked forecaster (typically 30–40 individuals, accounting for ties). The share is calculated only among forecasters who participated in both consecutive quarters. The line (right axis) indicates the number of such consecutive participants.

Figure A5: Relative Comparison of the Forecast Accuracy of the Top Performers



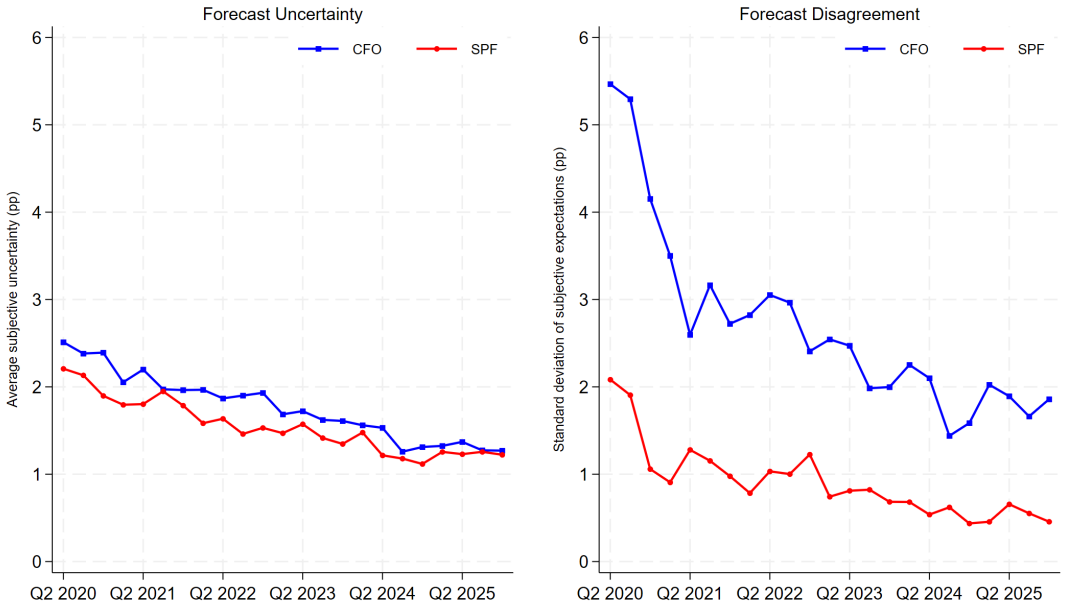
Notes: The plot on the left shows the mean absolute forecast errors of top performers and the rest of the panel for all quarters between 2021:Q2 and 2024:Q4. The right plot displays p-values from two-sample t-tests comparing mean absolute forecast errors of top-tier forecasters (identified in the previous quarter based on cumulative performance) versus the remaining panel, conducted separately for each quarter.

Figure A6: Overprecision between Surveys



Notes: This figure shows realized overprecision by target year using fourth-quarter surveys only. Each bar is the ratio of forecast RMSE to average subjective standard deviation; values above one (dashed line) indicate realized errors exceeding the uncertainty forecasters admitted (overprecise).

Figure A7: Forecaster Uncertainty and Disagreement (All Surveys)



Notes: The left-hand portion of this figure plots average subjective uncertainty within each quarter (in percentage points), measured as the mean across respondents of the standard deviation of each respondent's own probability distribution. The right-hand portion of this figure plots the cross-sectional standard deviation of subjective expectations within each quarter (in percentage points). Both charts show all survey quarters. Respondents who reported no probabilities are excluded (200 SPF records), and all retained responses sum to approximately 100 percent. The merged sample spans 2020Q2–2025Q4 (23 quarters), covering 649 SPF responses (mean 28 per quarter) and 8,301 CFO responses (mean 361 per quarter).

E Derivation of MSE Decomposition formula

$$
\begin{aligned}
\text{MSE}_t &= \frac{1}{N_t}\sum_{i=1}^{N_t}(y_t - \mathbb{E}_{it})^2 \\
&= \frac{1}{N_t}\sum_{i=1}^{N_t}\left[(y_t - \bar{\mathbb{E}}_t) + (\bar{\mathbb{E}}_t - \mathbb{E}_{it})\right]^2 \\
&= \frac{1}{N_t}\sum_{i=1}^{N_t}\left[(y_t - \bar{\mathbb{E}}_t)^2 + 2(y_t - \bar{\mathbb{E}}_t)(\bar{\mathbb{E}}_t - \mathbb{E}_{it}) + (\bar{\mathbb{E}}_t - \mathbb{E}_{it})^2\right] \\
&= (y_t - \bar{\mathbb{E}}_t)^2 + \underbrace{2(y_t - \bar{\mathbb{E}}_t)\frac{1}{N_t}\sum_{i=1}^{N_t}(\bar{\mathbb{E}}_t - \mathbb{E}_{it})}_{=0} + \frac{1}{N_t}\sum_{i=1}^{N_t}(\mathbb{E}_{it} - \bar{\mathbb{E}}_t)^2
\end{aligned}
$$

$$= \underbrace{(y_t - \bar{\mathbb{E}}_t)^2}_{\text{Bias}^2} + \underbrace{\frac{1}{N_t}\sum_{i=1}^{N_t}(\mathbb{E}_{it} - \bar{\mathbb{E}}_t)^2}_{=\text{Var}(\mathbb{E}_{it})}$$

References

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Brent Meyer, and Nicholas Parker (2022). "Surveying business uncertainty." *Journal of Econometrics*, 231(1), pp. 282–303. doi:https://doi.org/10.1016/j.jeconom.2020.03.021. URL https://www.sciencedirect.com/science/article/pii/S0304407620302785, annals Issue: Subjective Expectations Probabilities in Economics.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016). "Measuring economic policy uncertainty." *The Quarterly Journal of Economics*, 131(4), pp. 1593–1636. doi:10.1093/qje/qjw024. URL https://doi.org/10.1093/qje/qjw024.

Barrero, Jose Maria (2022). "The micro and macro of managerial beliefs." *Journal of Financial Economics*, 143(2), pp. 640–667. doi:https://doi.org/10.1016/j.jfineco.2021.06.007. URL https://www.sciencedirect.com/science/article/pii/S0304405X21002737.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992). "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy*, 100(5), pp. 992–1026. doi:10.1086/261849. URL https://doi.org/10.1086/261849.

Campbell, Sandy and Don A. Moore (2024). "Overprecision in the survey of professional forecasters." *Collabra: Psychology*, 10(1), p. 92,953. doi:10.1525/collabra.92953. URL https://doi.org/10.1525/collabra.92953.

Carroll, Christopher D. (2003). "Macroeconomic expectations of households and professional forecasters." *The Quarterly Journal of Economics*, 118(1), pp. 269–298. doi:10.1162/00335530360535207. URL https://doi.org/10.1162/00335530360535207.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia (2017). "Inflation expectations, learning, and supermarket prices: Evidence from survey experiments." *American Economic Journal: Macroeconomics*, 9(3), p. 1–35. doi:10.1257/mac.20150147. URL https://www.aeaweb.org/articles?id=10.1257/mac.20150147.

Clements, Michael P., Robert W. Rich, and Joseph S. Tracy (2023). "Chapter 3 - surveys of professionalsthe views expressed in the paper are those of the authors and do not necessarily reflect those of the federal reserve bank of cleveland, the federal reserve bank of dallas, or the federal reserve system.we are grateful to the handbook editors, as well as dean croushore and geoff kenny for helpful comments. we would also like to thank amber sherman for her assistance with the appendix table that accompanies this chapter." In Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw, editors, *Handbook of Economic Expectations*, pp. 71–106. Academic Press. doi:https://doi.org/10.1016/B978-0-12-822927-9.00009-4. URL https://www.sciencedirect.com/science/article/pii/B9780128229279000094.

Clements, Michael P. and Shixuan Wang (2026). "Do professionals' inflation forecasts incorporate

the beliefs of others? a functional data approach." *International Journal of Forecasting*. doi:https://doi.org/10.1016/j.ijforecast.2026.05.006. URL https://www.sciencedirect.com/science/article/pii/S0169207026000555.

Coibion, Olivier and Yuriy Gorodnichenko (2015). "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 105(8), p. 2644–78. doi:10.1257/aer.20110306. URL https://www.aeaweb.org/articles?id=10.1257/aer.20110306.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar (2018). "How do firms form their expectations? new survey evidence." *American Economic Review*, 108(9), p. 2671–2713. doi:10.1257/aer.20151299. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151299.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele (2020). "Inflation expectations and firm decisions: New causal evidence." *The Quarterly Journal of Economics*, 135(1), pp. 165–219. URL https://academic.oup.com/qje/article-abstract/135/1/165/5570592.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber (2022). "Monetary policy communications and their effects on household inflation expectations." *Journal of Political Economy*, 130(6), pp. 1537–1584. doi:10.1086/718982. URL https://doi.org/10.1086/718982.

Fried, Stephie, Tobin Graf, Simar Malhotra, and Sanjay R. Singh (2026). "What do financial officers predict for price growth?" FRBSF Economic Letter 2026-11, Federal Reserve Bank of San Francisco. URL https://www.frbsf.org/research-and-insights/publications/economic-letter/2026/05/what-do-financial-officers-predict-for-price-growth/.

Manski, Charles F. (2004). "Measuring expectations." *Econometrica*, 72(5), pp. 1329–1376. doi:10.1111/j.1468-0262.2004.00537.x. URL https://doi.org/10.1111/j.1468-0262.2004.00537.x.

Manski, Charles F. (2018). "Survey measurement of probabilistic macroeconomic expectations: Progress and promise." *NBER Macroeconomics Annual*, 32(1), pp. 411–471. doi:10.1086/696061. URL https://doi.org/10.1086/696061.

Moore, Don A. and Paul J. Healy (2008). "The trouble with overconfidence." *Psychological Review*, 115(2), pp. 502–517. doi:10.1037/0033-295X.115.2.502.

Rich, Robert W. and Joseph Tracy (2025). "All forecasters are not the same: Systematic patterns in predictive performance." *International Journal of Forecasting*. doi:https://doi.org/10.1016/j.ijforecast.2025.02.008. URL https://www.sciencedirect.com/science/article/pii/S0169207025000159.

Roth, Christopher and Johannes Wohlfart (2020). "How do expectations about the macroeconomy affect personal expectations and behavior?" *The Review of Economics and Statistics*, 102(4),

pp. 731–748. doi:10.1162/rest_a_00867. URL https://doi.org/10.1162/rest_a_00867.

Scharfstein, David S. and Jeremy C. Stein (1990). "Herd behavior and investment." *The American Economic Review*, 80(3), pp. 465–479. URL http://www.jstor.org/stable/2006678.

Song, Wenting and Samuel Stern (2024). "Firm inattention and the efficacy of monetary policy: A text-based approach." *The Review of Economic Studies*, 92(5), pp. 3438–3469. doi:10.1093/restud/rdae102. URL https://doi.org/10.1093/restud/rdae102.

Trueman, Brett (2015). "Analyst forecasts and herding behavior." *The Review of Financial Studies*, 7(1), pp. 97–124. doi:10.1093/rfs/7.1.97. URL https://doi.org/10.1093/rfs/7.1.97.

Weber, Michael, Bernardo Candia, Hassan Afrouzi, Tiziano Ropele, Rodrigo Lluberas, Serafin Frache, Brent Meyer, Saten Kumar, Yuriy Gorodnichenko, Dimitris Georgarakos, Olivier Coibion, Geoff Kenny, and Jorge Ponce (2025). "Tell me something i don't already know: Learning in low- and high-inflation settings." *Econometrica*, 93(1), pp. 229–264. doi:https://doi.org/10.3982/ECTA22764. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA22764.