# The Persistence of Miscalibration*

Michael Boutros, Itzhak Ben-David, John R. Graham,
Campbell R. Harvey, and John Payne†

November 2024

**Abstract**

We analyze a panel of over 28,400 S&P 500 return forecasts by CFOs to examine whether the extent of CFOs' miscalibration—providing forecast confidence intervals that are too narrow—decreases over time. We find no improvement with task repetition nor evidence of learning, that is, no improvement in response to past performance. Across CFOs, miscalibration appears to be a persistent personal trait. We find some evidence that the degree of miscalibration is related to birth cohort and stock market familiarity.

*We thank Itay Goldstein (editor), two anonymous referees, Neil Pearson (discussant), Manuel Adelino, Nick Bloom, Alon Brav, Elijah Broadbent, Steve Davis, Morad Elsaify, Bruno Feunou, Daniel Garrett, Daniel Kahneman, Florian Peters, and seminar participants at Duke University, the Developing and Using Business Expectations Data Conference at the University of Chicago, the Miami Behavioral Finance Conference, and the ITAM Finance Conference for comments.

†Boutros: University of Toronto. Ben-David: The Ohio State University and NBER. Graham: Duke University and NBER. Harvey (Corresponding Author): Duke University and NBER. Payne: Duke University.

Numerous psychological and economic studies show that executives, like the general population, exhibit a variety of behavioral biases. One common type of bias is miscalibration in belief formation—the systematic discrepancy between an individual's subjective confidence in their judgments or predictions and the objective accuracy of those judgments or predictions.[1] In the context of financial markets, Ben-David, Graham, and Harvey (2013) document severe miscalibration in stock market return forecasts by chief financial officers (CFOs), with only 36% of realized stock market returns falling within the 80% confidence intervals provided by these top executives.

The substantial miscalibration observed among top corporate decision-makers raises questions about the origins of miscalibration and, in turn, the potential for improving calibration. Prior literature proposes two main views on the nature of miscalibration. According to the first view, miscalibration may be eliminated with repetition, feedback, and learning.[2] The fact that miscalibration is commonly observed in practice, however, suggests that the necessary conditions for debiasing miscalibration are not met in real-life events (Thaler, 2000); for example, major events that generate learning opportunities may not occur or may not occur with sufficient repetition in an individual's lifetime. According to the second view, miscalibration may be a personal trait shaped by genetics or early life experiences that cannot be changed easily.[3] Evidence on the determinants of miscalibration and the impact of learning is mixed and comes mainly from laboratory experiments.[4] As Moore et al. (2015)

[1]Miscalibration, also known as overprecision, is one of three forms of overconfidence widely documented across individuals (Alpert and Raiffa, 1982; Soll and Klayman, 2004; Moore and Healy, 2008; Mannes and Moore, 2013; Moore and Schatz, 2017; Moore and Dev, 2018; Campbell and Moore, 2024). Overconfidence is "excessive faith that [the agent] knows the truth," while overprecision is "excessive confidence that one's beliefs are accurate" (Moore and Swift, 2011; Moore and Schatz, 2017). For example, laboratory examples of overprecision include estimating the number of eggs produced in a calendar year (Alpert and Raiffa, 1982) or the year of the first hot air balloon flight (Soll and Klayman, 2004).

[2]Thaler (1987, 1994, 2000); Moore, Tenney, and Haran (2015). Evidence regarding the impact of learning on behavioral biases in other finance domains, such as the behavior of individual and institutional investors, is also mixed (Feng and Seasholes, 2005; Kaustia and Knüpfer, 2008; Chiang, Hirshleifer, Qian, and Sherman, 2011).

[3]Cesarini, Lichtenstein, Johannesson, and Wallace (2009); Johnson and Fowler (2011); Malmendier and Nagel (2011); Malmendier, Tate, and Yan (2011)

[4]Lichtenstein and Fischhoff (1980); Benson and Önkal (1993); Subbotin (1996); Baranski and Petrusic (1999); Sieck and Arkes (2005); González-Vallejo and Bonham (2007); Rakow, Harvey, and Finer (2003); McKenzie, Liersch, and Yaniv (2008).

summarizes the state of the literature: "overprecision remains an important phenomenon in search of a full explanation."

In this paper, we present new evidence consistent with the view that miscalibration is a persistent personal trait that does not adjust easily with repetition and learning opportunities. Using 22 years of S&P 500 return forecasts provided by CFOs, along with their reported confidence intervals, we show a persistent pattern of miscalibration in this repeated forecasting task. We observe no aggregate improvement in calibration levels over our 22-year period. Moreover, we find no improvement within individual forecasters as they produce more forecasts, nor do they materially adjust their confidence interval in response to past realizations; that is, we find no evidence of learning. Instead, most of the variation in miscalibration is explained by time-invariant person-fixed effects. Among CFO characteristics, CFOs' birth cohorts and familiarity with the stock market are correlated with CFOs' average miscalibration, while corporate characteristics like firm size and industry are not. Taken together, our results suggest that miscalibration is a persistent personal trait that is not sensitive to the corporate environment and is potentially related to innate traits and early life experiences.

Our study is based on a unique dataset of forecasts by CFOs. Specifically, in each quarter from 2001Q2 to 2023Q1, the Duke-CFO Survey asks CFOs to report their predictions of one-year returns on the S&P 500. CFOs are also asked to provide 80% confidence intervals (CIs) around their forecasts. These data allow us to impute CFOs' beliefs about return volatility. Overall, the sample comprises over 28,400 predictions of one-year returns on the S&P 500 by more than 6,700 CFOs over the 2001 to 2023 period (86 quarters), with over 10% of the CFOs making at least eight forecasts.

Our study consists of three main parts. In the first part, we examine whether CFOs' calibrations have improved over the sample period. Since the early 2000s, the academic world has uncovered extensive evidence of behavioral biases among executives and the impact of such biases on corporate decisions. General understanding of these biases has increased with Nobel Prizes awarded to Kahneman (2002), Shiller (2013), and Thaler (2017) for significant contributions in this domain. One might expect greater awareness of these psychological biases to improve CFOs' calibrations. Contrary to this expectation, however, we find that the calibration levels recorded in our quarterly surveys did not improve over the 22-year sample. The average width of CFOs' confidence intervals was approximately 14% throughout the period, whereas a well-calibrated 80% CI based on historical data should be three times wider. Furthermore, we find that only 29.7% of realized S&P 500 returns fall within the CFOs' forecasted 80% CIs, that is, the CFOs' "hit rate" is less than 30%. Our results show that the average CFO hit rate has remained relatively constant throughout the sample period. The earlier findings in Ben-David et al. (2013), therefore, continue to hold more than a decade out-of-sample.

In the second part of our study, we examine how CFO miscalibration varies with repetition, as well as how it responds to past performance. The strength of our analysis lies in a large panel of forecasts for thousands of forecasters. Since the same CFO respond multiple times to our quarterly surveys, we can estimate the importance of CFOs' baseline miscalibration levels and whether their miscalibration evolves over time with repetition.

The baseline level of CFOs' miscalibration level varies a lot across CFOs and is highly persistent for each specific CFO, akin to a personal attribute. CFO fixed effects explain 59.6% of the variation (as measured by adjusted $R^2$) in CFOs' reported CIs. In comparison, calendar time fixed effects explain only 3.3% of the variation in CIs. Conversely, the likelihood of hitting the CI—which depends on CFOs' forecasts of market returns—is determined primarily by time fixed effects: CFO fixed effects explain 10.4% of the variation of hitting the CI, while calendar time fixed effects, which capture unexpected stock market movements, explain 33.5% of this variation.

We observe little evidence that the extent of miscalibration decreases with repeated learning opportunities. In subsequent forecasts, CFOs' forecasts show no material change in CIs or hit rates relative to their initial projections. In addition, when we employ a model of Bayesian learning to predict empirical patterns indicative of CFOs learning from their own past forecast accuracy (i.e., whether they hit or miss their past CIs), we find that contrary to our model's predictions, there is no evidence of CFOs adjusting their CIs based on the accuracy of their prior forecasts.

In the third part of the paper, we explore the factors correlated with CFO miscalibration levels. Given that CFO fixed effects best explain miscalibration, we focus on CFOs' personal and organizational characteristics as potential sources of variation that can explain this bias.

Our results are most consistent with the view that early life experiences drive miscalibration. We present evidence that CFOs' birth year, not their age, has some predictive ability for miscalibration (result akin to Malmendier and Nagel, 2011). Furthermore, while CFOs in large firms and in certain industries exhibit improved calibration, this result may well stem from selection bias rather than the direct influence of the corporate environment on miscalibration, given that we do not detect a material change in their miscalibration as we track CFOs across employment at firms of different sizes and industries.

Interestingly, we also document an inverse U-shaped relationship between CFOs' self-reported familiarity with the stock market and the degree of their miscalibration. Specifically, CFOs at the ends of the miscalibration spectrum—either most or least miscalibrated—report relatively low familiarity with the stock market. In contrast, those in the middle of the miscalibration spectrum say they are familiar with the stock market. The fact that CFOs who are less familiar with the stock market are more miscalibrated is reminiscent of the Dunning-Kruger effect (Kruger and Dunning, 1999). On the other hand, CFOs who perceive themselves to be experts appreciate the uncertainty inherent in the stock market, which decreases the extent of miscalibration, but at the same time, are more certain about their beliefs, which increases the extent of miscalibration. We provide evidence that the former effect dominates for moderate levels of expertise, while the latter dominates for high levels of expertise.

Given that our study is based on a unique, large-scale, decades-long survey of real-world financial executives, our approach has several differences relative to a laboratory setting, which have implications for our analysis. First, our respondents are not typical student lab participants or novices but rather are financial executives who are generally well-versed in the stock market. At the same time, CFOs respond to the survey voluntarily, potentially causing selection effects. We address potential selection effects in the analysis. Second, the object of interest is not abstract or hypothetical but rather a real-life belief with material relevance. Third, the feedback that CFOs receive is a public signal, namely realized S&P 500 returns, which are readily and simultaneously available to all participants, who may differ in their levels of interests and familiarity. Fourth, our study is not a one-shot experiment but rather features a large panel of repeat respondents who answer an identical question each quarter. CFOs who participate in several surveys are naturally exposed to financial news and have the opportunity to learn from past mistakes at a high frequency. Our survey should be viewed as a way to measure CFOs' beliefs over time rather than an intervention attempting to improve their calibration.

In summary, our study provides novel evidence on the persistence of miscalibration among CFOs. This evidence has direct implications for the study of decision-making with overconfident executives. Malmendier and Tate (2015) survey a large literature in corporate finance that documents the impact of managerial overconfidence in areas such as governance, innovation, dividend policy, investment, and acquisitions.[5] Several studies focus specifically on the implications of biased beliefs (Greenwood and Shleifer, 2014; Pflueger, Siriwardane, and Sunderam, 2020). Our evidence suggests that time and experience do not mitigate miscalibration among top executives. These findings underscore the challenge of correcting miscalibration and emphasize the need for further research on effective interventions to address cognitive biases in professional decision-making settings.

[5] Malmendier and Tate (2005, 2008); Goel and Thakor (2008); Galasso and Simcoe (2011); Ben-David et al. (2013); Deshmukh, Goel, and Howe (2013); Hirshleifer, Low, and Teoh (2012); Barrero (2022); Graham (2022).

# 1 Data

## 1.1 Survey Sample

We analyze a panel of stock market predictions made by financial executives in the Duke-CFO survey. The survey is delivered electronically to senior financial executives each quarter.[6] The survey contains questions that appear every quarter and topical questions that poll CFOs on important events related to current economic and geopolitical conditions. This dataset has been used in several prior academic studies.[7]

Our dataset spans 86 quarterly surveys from 2001Q2 to 2023Q1, comprising over 28,400 individual observations. This is a substantial expansion from the 40 surveys and 13,000 observations used by Ben-David et al. (2013), covering 2001Q2 to 2011Q2. We are able to identify 6,760 distinct CFOs, many of whom respond to the survey multiple times. We use this new dimension of the dataset to study the evolution of their confidence intervals over time.

Figure 1 provides survey participation statistics. Panel (a) presents the distribution of the number of forecasts collected in each survey date. In most quarterly surveys, the number of respondents varies between 200 and 500. In the early period (pre-2005Q1), we could not identify CFOs; hence, these survey waves are excluded from our analysis of the evolution of individual responses. Panel (b) presents a histogram of the number of forecasts by the same CFO. There is a wide variation in the number of surveys in which each CFO participates. Just over half of the CFOs drop out after making a single forecast. Approximately 1,000 executives respond to the survey exactly twice. More than 450 respondents respond at least

[6]Through early 2002, the survey was delivered via fax. Since mid-2002, invitations to participate in surveys have been emailed, and CFOs have entered their responses through a web portal. Starting in June 2020, the survey has been administered jointly by Duke University and the Federal Reserve Banks of Atlanta and Richmond. In two surveys during the transition period (2019Q1 and 2020Q2), we did not include the S&P 500 forecast question; in one survey (2020Q1), the question was sent only to a subset of participants.

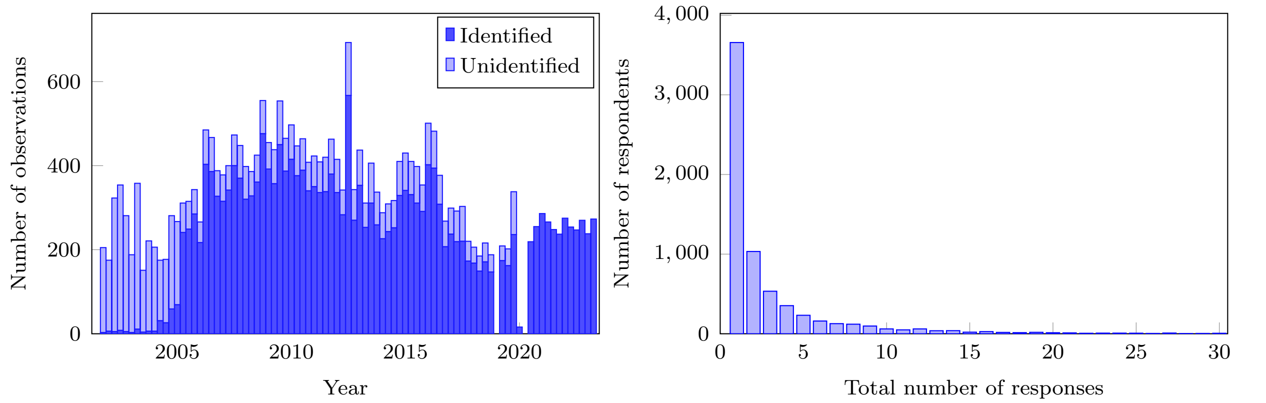
[7]For example, Ben-David et al. (2013) study miscalibration patterns for CFOs, Greenwood and Shleifer (2014) analyze a long time series of investor expectations and their relationship to expected returns in standard finance models, and Gennaioli, Ma, and Shleifer (2016) document that CFO forecasts accurately anticipate future investment.

**Figure 1. Count of Responses Over Time and per Respondent**

This figure presents descriptive statistics regarding our sample. Panel (a) shows the number of identified and unidentified responses to the survey by survey date. Panel (b) presents a histogram of the total number of responses provided by identified respondents.

(a) Number of responses, per survey

(b) Histogram of number of forecasts per respondent




10 times, and more than 100 CFOs respond to at least 20 surveys.

Following the recommendation by List (2007), we benchmark our sample against Compustat companies across four different demographic characteristics (see Appendix A).[8] Because our sample includes both private and public companies, we provide separate comparisons for each type of firm relative to the public firm data from Compustat. Including private and public companies in our sample enhances its representativeness of the broader U.S. economy. Not surprisingly, our full sample (private and public), has a higher proportion of smaller companies (21.6%) compared to the Compustat sample (12.5%). Isolating the public firms in our sample indicates that our sample has larger public firms than Compustat, with 16.2% having sales above $10 billion compared to 7.5% in Compustat. This is inconsistent with the notion that busy CFOs do not participate in our surveys. Our sample has a smaller proportion of technology companies and a larger proportion of service/consulting companies compared to Compustat. Overall, while there are some differences between our sample and Compustat, the differences are unremarkable and do not suggest that sample composition drives our results.

[8]See Ben-David et al. (2013), Graham, Harvey, and Puri (2013), and Graham (2022) for additional analysis on the representativeness of the Duke-CFO survey sample.

## 1.2 Measuring Miscalibration

The main questions of interest pertain to CFOs' beliefs about the future one-year S&P 500 return:

*Over the next year, I expect the annual S&P 500 return will be:*

*– There is a 1-in-10 chance the actual return will be less than __%.*

*– I expect the return to be: __%.*

*– There is a 1-in-10 chance the actual return will be greater than __%.*

Following the standard practice in the literature, these three questions elicit a point estimate for the mean expected return and an 80% CI around it (Soll, Milkman, and Payne, 2015a,b; Bloom, Davis, Foster, Lucking, Ohlmacher, and Saporta-Eksten, 2017; Fermand, Kuhnen, Li, and Ben-David, 2024). Based on the 80% CI provided by a forecaster, one can impute the standard deviation of the stock market return that the forecaster expects.[9] Forecasters were not reminded of their previous survey responses.

In our analysis, we rely on two key measures for CFOs' miscalibration:

1. **Ex-ante miscalibration measure: Width of the confidence interval.** Our ex-ante miscalibration measure is the width of the CI that a CFO provides, measured in percentage points. In our sample, the average CI is about 14%. As a benchmark, the $10^{\text{th}}$ and $90^{\text{th}}$ percentiles of 12-month S&P 500 returns over the 1950 to 2023 period are $-12.5\%$ and 28.1%, respectively. Historical returns, therefore, suggest that a well-calibrated 80% confidence interval should be about 43% per year.

2. **Ex-post miscalibration measure: Hitting the confidence interval.** The ex-post miscalibration measure is a binary variable indicating whether the realized S&P 500 return fell within the CI provided 12 months earlier by the CFO. Well-

[9] Keefer and Bodily (1983) show that the following method is preferred for calculating the imputed standard deviation of a continuous random variable, given the $10^{\text{th}}$ and $90^{\text{th}}$ percentiles: $\sigma = \frac{P_{90}-P_{10}}{2.65}$. To ensure that outliers do not drive our results, we exclude observations with CIs in the top and bottom percentiles.

calibrated forecasters should hit the 80% CI 80% of the time. A lower hit rate indicates greater miscalibration. Over the full sample period, the average hit rate is 29.7%.

## 1.3 CFO Characteristics

We collect additional information about respondents' demographics, employment, and stock market familiarity. The summary statistics are provided in Appendix Table A.1.

**Age and birth cohort.** In two surveys (2008Q1 and 2014Q4), CFOs were asked their age. We propagate CFO birth years and ages across other surveys to which they respond.

**Firm size.** In every survey, we ask CFOs to indicate the scale of their firm's revenues, which we use as a measure of firm size.

**Industry.** In every survey, we ask CFOs to indicate their industries: (1) Retail/Wholesale, (2) Mining/Construction, (3) Manufacturing, (4) Transportation/Energy, (5) Communications/Media, (6) Tech (Software/Biotech), (7) Banking/Finance/Insurance, (8) Service/Consulting, (9) Healthcare/Pharmaceutical, or (10) Other.

**Stock market familiarity.** Since 2020Q1, the survey also includes a question that allows us to measure CFOs' perceived familiarity with the stock market. Specifically, CFOs are asked to choose between the following options: (1) I do not follow the stock market, (2) I look at the stock market occasionally, (3) I follow the stock market closely (not for work), (4) I follow the stock market closely (for work), or (5) Other.

Based on these responses, we construct an indicator of familiarity with the stock market, which we interpret as a measure of a CFO's expertise with respect to the forecasting task. This variable is set to one for CFOs who say they follow the stock market (responses (3) and (4)) and zero otherwise. For each CFO, we calculate their average stock market familiarity for all surveys in which they participate, and propagate this average familiarity for each CFO to all of their responses. Based on this measure, 81% of CFOs are familiar with the stock market, and the remaining 19% report looking at the stock market occasionally (response (2)).

## 1.4 Summary Statistics

Table 1 presents some summary statistics for the sample used in the study. In total, 6,760 CFOs made 28,413 forecasts. The variable *forecast number* corresponds to the forecast number for a specific forecaster. The average forecast number in our sample is 5.2, which speaks to the richness of our panel dataset.

As discussed in the next section, miscalibration is pervasive throughout our sample. Compared to the width of a properly calibrated 80%-confidence interval (43%), the average width of CIs is only 14.1%. The median width is 10%, and even the $90^{\text{th}}$ percentile is only 30%. As such, the hit rate, captured by the indicator variable I(Hit CI), is only 29.7%.

Based on the categorical firm size variable, the average firm size in our sample is \$100m–\$500m, and there is coverage across all firm sizes. At the $25^{\text{th}}$ percentile, firm size is less than \$25m; firm size at the $75^{\text{th}}$ percentile is \$500m–\$1bn. Birth year and age are available for a subsample of CFOs. On average, CFOs are 53.7 years old, with a standard deviation of eight years. This variation allows us to use these company and demographic variables to study "fixed effects" of CFOs on their forecasts.

## 1.5 Concerns with Self-Reported Beliefs and Selection

Self-reported beliefs about stock market returns may suffer from two distortions: survey respondents may be unfamiliar with market returns, or they may not report their true beliefs.[10] The overwhelming majority of respondents are familiar with the stock market, which is unsurprising given their employment as senior financial executives. As discussed

[10] Other papers that use the CFO survey dataset discuss these distortions in the context of their analyses (e.g., Greenwood and Shleifer, 2014, Section 2).

**Table 1. Summary Statistics**

This table presents summary statistics for the sample used in the study. The sample covers 86 survey quarters from 2001Q2 to 2023Q1. See Section 1 for details on the sample construction.

| Variable | N | Mean | StDev | Min | p10 | p25 | p50 | p75 | p90 | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| Upper confidence bound (%) | 28,413 | 10.75 | 7.09 | −40 | 5 | 6 | 10 | 15 | 20 | 115 |
| Expected return (%) | 28,413 | 5.16 | 5.45 | −50 | 0 | 3 | 5 | 8 | 10 | 105 |
| Lower confidence bound (%) | 28,413 | −3.34 | 9.32 | −60 | −15 | −10 | 0 | 2 | 5 | 97 |
| Confidence interval (CI) (%) | 28,413 | 14.09 | 11.61 | 1 | 4 | 6 | 10 | 20 | 30 | 150 |
| $\Delta\text{CI}_{q-4\to q}$ (%) | 5,835 | 0.35 | 10.56 | −95 | −10 | −5 | 0 | 5 | 11 | 93.4 |
| I(Hit CI) | 27,385 | 0.30 | 0.46 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| I(Miss CI) | 27,385 | 0.70 | 0.46 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Forecast number | 20,921 | 5.16 | 6.10 | 1 | 1 | 1 | 3 | 7 | 13 | 55 |
| log(Forecast number) | 20,921 | 0.49 | 0.43 | 0.0 | 0.0 | 0.0 | 0.5 | 0.8 | 1.1 | 1.7 |
| Birth year | 5,005 | 1958.62 | 8.33 | 1932 | 1948 | 1954 | 1960 | 1964 | 1970 | 1983 |
| Age | 5,005 | 53.74 | 8.63 | 29 | 43 | 48 | 54 | 59 | 65 | 88 |
| Firm size | 27,950 | 2.96 | 1.68 | 1 | 1 | 2 | 3 | 4 | 5 | 7 |
| I(S&P familiarity) | 4,355 | 0.81 | 0.33 | 0.0 | 0.1 | 0.7 | 1.0 | 1.0 | 1.0 | 1.0 |

later in Section 2.1, the average expected return provided by CFOs is very close to the average S&P 500 annual return. While this similarity does not necessarily imply that all CFOs are good forecasters, it does imply that they understand the survey question and take it seriously. For these reasons, we believe that the responses elicited by the survey are a fair representation of respondents' beliefs.

There may also be a concern that CFOs who respond to the survey are the most miscalibrated. This may be the case if, for example, CFOs with the lowest value of time are the ones who respond to the survey and the same CFOs are the most miscalibrated. It may also be the case that CFOs who are less miscalibrated drop out of the survey. Instead, we find some evidence in the raw sample that CFOs who are *more* miscalibrated are more likely to drop out of the sample. Failing to control for this composition effect would bias the results toward appearing as though overall miscalibration improved with repetition. In Appendix D, we show that controlling for CFO fixed effects can mitigate these composition effects. We employ these controls in each of our regressions.

**Table 2. Moments of CFO Beliefs and Realized Returns**

CFOs' beliefs over the return distribution are calculated using their forecasts of the S&P 500 return. The mean is the average forecast of future S&P 500 returns and the standard deviation is the average imputed belief of volatility (measured using the formula in footnote 9). To compare to the historical data, we calculate annual returns in each quarter from the listed date to 2023Q1, the last quarter in our sample. The last column corresponds to our sample period from 2001Q2 to 2023Q1. The standard deviation ratio, $\lambda$, is the ratio of the CFOs' belief of the standard deviation, 5.3%, and the historical standard deviation.

| | | Historical returns (Date to 2023Q1) | | | | |
|---|---|---|---|---|---|---|
| | CFO beliefs | 1960Q1 | 1970Q1 | 1980Q1 | 1990Q1 | 2001Q2 |
| Mean | 5.2% | 8.2% | 8.6% | 10.1% | 9.1% | 6.4% |
| Standard deviation | 5.3% | 16.6% | 17.2% | 17.0% | 16.8% | 17.6% |
| Ratio of Beliefs to Historical | — | 0.320 | 0.309 | 0.313 | 0.316 | 0.302 |

# 2 Historical Return Data and Miscalibration Over Time

In this section, we compare CFOs' reported volatility beliefs (expressed via their confidence intervals) to the S&P 500's historical realized volatility and explore the evolution of the average level of CFOs' miscalibration over our 22-year sample period.

## 2.1 CFO Beliefs and Historical Return Data

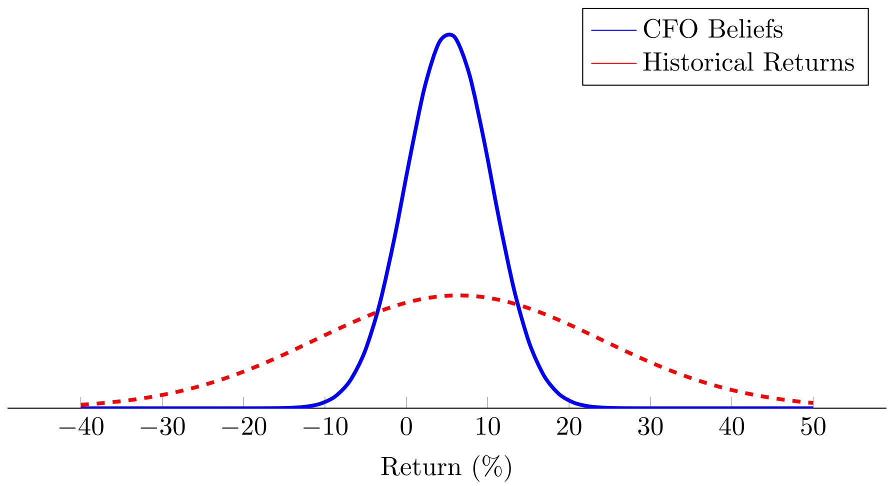
We contrast CFOs' stock market forecasts with the market's historical performance to understand the reasonableness of their forecasts and assess the degree of their miscalibration. In Table 2, the first column presents the average forecasted return and the average imputed standard deviation for each forecast. The remaining columns contain the mean and standard deviation of historical one-year S&P 500 return distributions over different periods.

The last column in Table 2 presents the average and standard deviation of quarterly annual returns over our survey sample period from 2001Q2 to 2023Q1. Over this period, the average annual return was 6.4%, whereas the average of CFOs' beliefs over these returns was 5.2%.

Table 2 demonstrates the miscalibration in our sample. The standard deviation of the S&P 500's realized annual returns was 17.6% over the sample period, more than three times

**Figure 2. Distributions of CFO Forecasts and Historical Returns**

This figure presents the difference between the historical distribution of realized S&P 500 returns and CFOs' beliefs over the return distribution imputed from their forecasts (see Table 2). Though the data exhibit excess kurtosis and are not normally distributed, we illustrate these two distributions as normal distributions with mean and standard deviation. The historical returns distribution is parameterized by the mean and standard deviation of the returns distribution over the sample period, which are 6.4% and 17.6%, respectively. The mean for the CFO beliefs distribution, 5.2%, is the average forecast of future S&P 500 returns, and the standard deviation, 5.3%, is the average of the imputed standard deviations, measured using individual CFOs' 80% CIs.




larger than the average CFO belief of the standard deviation, 5.3%. The third row of the table calculates the ratio between CFOs' beliefs over the standard deviation and the historical standard deviation. Overall, forecasters' beliefs over the standard deviation are approximately one-third of the sample standard deviation.[11]

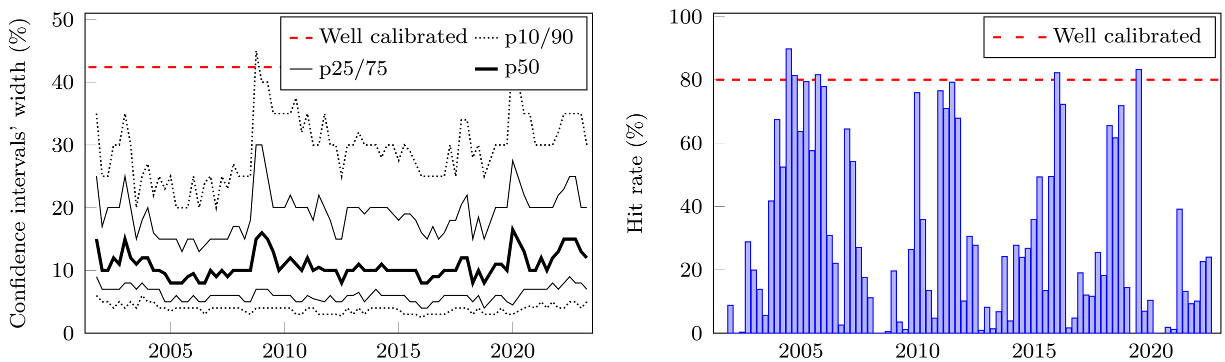
Thus, CFOs' beliefs over standard deviation (and variance) are poorly calibrated. Figure 2 illustrates the degree of miscalibration in our sample by plotting two normal distributions parameterized by the means and standard deviations of CFO beliefs and realized returns over our sample period. Even if forecasters can correctly gauge the (unconditional) mean of the return distribution, their belief over the 80% CI is far too narrow relative to the true distribution, and as a result, realized returns frequently fall outside this interval.

[11] Barrero (2022) estimates a dynamic model that allows managers to have miscalibrated beliefs and finds that managers underestimate the true volatility of the stochastic process by approximately 46%, which is in line with our findings.

**Figure 3. CFO Miscalibration Over Time**

This figure presents the average CI width for each survey (Panel (a)) and the average hit rate (Panel (b)). The panels show the theoretical values of well-calibrated CI widths and hit rates, respectively, in red dashed horizontal lines.

(a) Confidence intervals over time (b) Hit rate over time




## 2.2 Miscalibration Over Time

Next, we explore whether the average level of CFO miscalibration improved over the sample period. Given the increasingly wide discussion in academia and public media about the psychological effects on decision making, we hypothesize that miscalibration levels should be improved as CFOs are presumably more aware of these effects.

Figure 3, Panel (a) presents the distribution of CI widths over time. Over the sample period, the miscalibration measures show no obvious trends; the median CI is around 10% throughout the sample period. Instead, both the hit rate and the width of the CI appear to be driven by market movements and uncertainty. In response to economic stress, CFOs provide wider CIs.

Panel (b) shows the percentage of responses in each survey that hit the forecast interval. Over the full sample period, CFOs' hit rate reaches 80% on only a handful of survey dates. In most quarters, hit rates are materially lower. The sample average hit rate is 29.7%. Realized market volatility impacts CFOs' hit rates dramatically, as forecasts made ahead of large market swings are much more likely to result in missing the confidence interval than hitting it. For example, during the Global Financial Crisis of 2008 to 2009, CFOs' hit rates were zero or close to zero in several quarters.

**Table 3. CFO Miscalibration Over Time**

This table presents estimates from ordinary least squares regressions of average miscalibration measures regressed on the survey's date (measured as a calendar year and a fraction for mid-year surveys). Miscalibration measures are averaged for each survey. The average confidence interval is the miscalibration measure used in Columns (1)–(2). *Year* is the survey date measured in years, e.g., 2015.75 is the survey conducted in 2015Q4. The miscalibration measure used in Columns (3)–(4) is the hit rate of S&P 500 into CFOs' confidence intervals. Columns (1) and (3) use the entire sample to calculate miscalibration measures. Columns (2) and (4) use only the first forecast of each CFO when calculating the miscalibration measures. Before 2005Q2, only a few CFOs were identified, the analysis in Columns (3)–(4), which rely on first forecasts only, is restricted to the period of 2005Q2 onward. 2020Q1 is excluded as it only contains 16 observations. Standard errors are adjusted to autocorrelation (four-quarter lags) to account for the difference between the frequency of the forecasts (quarterly) and their horizon (annual), following Newey and West (1986). ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively, assuming a single test.

| Dependent variable: | Avg CI (%) | | Hit rate (%) | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Year | 0.072 | 0.126 | −0.535 | −1.047 |
| | (0.069) | (0.088) | (0.761) | (1.050) |
| Sample | All | 1st forecasts | All | 1st forecasts |
| Sample start date | 2001Q4 | 2005Q2 | 2001Q4 | 2005Q2 |
| Observations | 85 | 70 | 81 | 66 |
| $R^2$ | 0.034 | 0.063 | 0.011 | 0.033 |

Table 3 uses regressions to formally assess the time trend in the average miscalibration measures over time. The sample is a time series of the average miscalibration measures—i.e., one observation per survey date. The independent variable is the survey date, measured in years (e.g., Year = 2015.75 represents the survey conducted in 2015Q4).

The sample used in the results presented in Columns (1) and (3) of Table 3 includes all of the forecasts in our sample. These results generally do not reveal a statistically significant time trend in either miscalibration measure (average CI or hit rate). Since most of the forecasts are made by CFOs who respond to the survey multiple times, one may be concerned that forecaster learning or time-varying sample composition distort the time trends of the miscalibration measures. To mitigate these effects, we rerun the analysis including only the first forecast of each CFO (limited to respondents we can identify). The results in Columns (2) and (4) show no statistically significant time trend in either miscalibration variable. Overall, the lack of improvement in the aggregate measures of miscalibration are a first indication that CFOs, as a whole, are not learning over time.

# 3 The Impact of Task Repetition and Learning Opportunities

In this section, we explore whether CFO miscalibration improves with repetition. In other words, we explore whether miscalibration is a time-invariant characteristic or whether it evolves dynamically.

Our analysis is performed in three steps. First, we assess the baseline contribution of individual-level fixed effects in explaining the observed CI (Section 3.1). Second, we explore whether repetition of the forecasting task is associated with calibration improvement (Section 3.2). Finally, we test a learning model that predicts that CIs in repeat forecasts should expand and contract in response to the performance of past forecasts (Section 3.3).

## 3.1 Baseline: CFO Fixed Effects

We begin by measuring the explanatory power of individual-level fixed effects. In Table 4, we present the results of regressions of the miscalibration proxies on time-fixed effects and CFO-fixed effects.

The baseline results in this table highlight the main factors driving the two miscalibration measures: CI and hit rate. Columns (1) and (2) reveal that CFO fixed effects are the primary determinants of CI, explaining much of the variation with an $\mathrm{R}^2$ of 0.57 in Column (2). By contrast, time-fixed effects account for only a small portion of CI variation, with an $\mathrm{R}^2$ of 0.03 in Column (1). CIs are generally wider during economic stress, as seen in Panel (a) of Figure 3, but overall explain a small part of the variation in CI. Including both fixed effects marginally increases the explanatory power, with an $\mathrm{R}^2$ rising to 0.60 in Column (3).

Conversely, CFOs' hit rates are mainly driven by time-fixed effects: the $\mathrm{R}^2$ for time-fixed effects alone is 0.34 in Column (6), while CFO fixed effects alone yield an $\mathrm{R}^2$ of 0.10 in

**Table 4. CFO Fixed Effects and Repeat Forecasts**

This table shows estimates from ordinary least squares regressions of miscalibration measures on the forecast ordinal number for the CFO who makes the forecast. Standard errors are double-clustered by the survey's calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent Var.: | CI (%) | | | | | I(Hit CI) ×100 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Forecast # | | | | 0.019 | | | | | 0.220 | |
| | | | | (0.037) | | | | | (0.132) | |
| log(Forecast #) | | | | | 1.061** | | | | | 4.374** |
| | | | | | (0.465) | | | | | (1.827) |
| Survey date FE | Yes | No | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Forecaster FE | No | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes |
| Observations | 20,921 | 17,264 | 17,264 | 17,264 | 17,264 | 19,893 | 16,304 | 16,304 | 16,304 | 16,304 |
| Adjusted $R^2$ | 0.033 | 0.573 | 0.596 | 0.596 | 0.596 | 0.335 | 0.104 | 0.444 | 0.444 | 0.444 |

Column (7). This distinction arises because hit rates depend both on CI width at the time of forecasting and, much more heavily, on S&P 500 price movement in the subsequent year. Still, incorporating CFO fixed effects enhances the explanatory power for hit rates, boosting the $R^2$ to 0.44 in Column (8).

These results motivate our focus on the variation in confidence intervals as a proxy for miscalibration. Specifically, to study the determinants of miscalibration, we should concentrate on understanding the individual-level fixed effects that drive CI variation, rather than focusing on CFOs' hit rates: why some CFOs provide wider CIs than others.

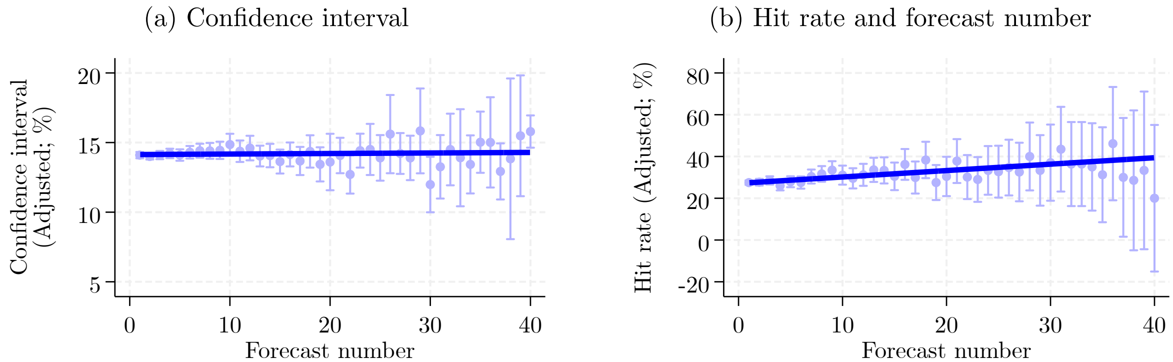
## 3.2 Experience: Improving with Repetition

Next, we explore whether repeat forecasts help explain CIs and hit rates. In Table 4, Column (3), we regress CIs on CFOs' forecast number and calendar quarter and CFO fixed effects. In Column (4), we replace the forecast number with the log (base 10) of the forecast number due to the skewness of this variable and for ease of interpretation. We repeat the process in Columns (7) and (8) for the hitting indicator as the dependent variable.

The results show that *within*-CFO (that is, net of CFO fixed effects), neither CIs nor

**Figure 4. Miscalibration Measures and Forecast Number, Within CFOs**

This figure presents the relation between the average CI (%; Panel (a)) and the average hit rate (%; Panel (b)) as a function of CFOs' forecast number. The generalized binscatter plots are created using the methodology and programs described in Cattaneo et al. (2024). $y$-axis variables are adjusted for calendar time and CFO fixed effects (the sample average was added back). The sample includes surveys from 2005Q2 to 2019Q4 and 2020Q2 to 2023Q2.




hit rates materially move with repetition. Our miscalibration measures are not correlated with the forecast number that CFOs make (Columns (3) and (7)), i.e., there is no change in miscalibration as CFOs make more forecasts. The correlation becomes statistically significant at the 5% level only when considering the log of forecast number. The economic effect of these results, however, is small. Column (4) shows that increasing the number of forecasts by an order of magnitude (e.g., from 1 to 10 or from 10 to 100) increases CIs by 1.0pp (percentage point) and the hit rate by 4.4pp. At this rate of improvement, the typical CFO in our sample would need many millennia of quarterly surveys to become well-calibrated.

Similar results can be seen in Figure 4. This figure plots the miscalibration measures per forecast number. The miscalibration measures are adjusted for calendar time and CFO fixed effects and calculated as regression residuals. The sample mean is added back. This figure should, therefore, be interpreted as showing the evolution of CFOs' miscalibration measures within-CFO, controlling for calendar fixed effects. In the left panel, variation in the confidence interval spans only a few percentage points, and the confidence bands on each estimate widen as the number of forecasts increases due to fewer observations. The figure shows no material evolution of CFO miscalibration with the number of forecasts. In the right panel, the hit rate marginally improves as the CI widens, but again, the confidence bands around each estimate widen due to fewer observations. In sum, CFOs begin with a poorly miscalibrated confidence interval that does not improve materially with repetition.

## 3.3 Learning: Improving Based on Past Outcomes

Next, we test a learning hypothesis, that is, whether CFOs adjust their CIs in response to past performance. The idea is that even if the base level of CFOs' CIs is too low in the first forecasts, they may adjust their CIs in later forecasts as they realize that their actual hit rate is too low relative to 80% benchmark.

The learning hypothesis proposes that forecasters gain experience and, in turn, learn over time. Learning could occur through several channels. For example, forecasters may pay attention to the stock market and update their perception of volatility, and/or they may learn from their earlier mistakes and update their CIs accordingly.

To guide our empirical analysis, we develop a model that predicts how Bayesian forecasters should update their predictive CIs by learning from past hits and misses of past CIs. In the model, which is detailed in Appendix B, agents use Bayes' rule to update their beliefs about the unknown variance of stock market returns by observing whether realized stock returns fall within their previously stated CIs.[12]

The model predicts that forecasters dynamically update their CIs based on feedback from their past performance. If the realized return falls within the CI, the CI may have been too wide, in which case the forecaster should narrow the interval in the subsequent forecasting round. However, if the realized return misses the CI, the forecaster should update her belief about volatility upward and widen the CI in the following forecasting round. This empirical

[12] A growing literature documents heterogeneity in the formation and evolution of beliefs and the impact of this heterogeneity on outcomes. See, for example, Gervais and Odean (2001), Kuhnen (2014), Kuchler and Zafar (2019), Meeuwis, Parker, Schoar, and Simester (2022), Giglio, Maggiori, Stroebel, and Utkus (2021), Martin and Papadimitriou (2022), and Fermand et al. (2024).

**Table 5. Tests of the Bayesian Learning Model**

This table shows estimates from ordinary least squares regressions of $\Delta\text{CI}_{q-4\to q}$, the change in CFOs' CIs from one forecast to the next (with a four-quarter spacing). $\text{I(Miss CI)}_{i,q-4}$ is a lagged indicator of whether the S&P 500 return missed the previous CI. $\text{CI}_{q-4}$ is the previous CI. Standard errors are double-clustered by survey calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | $\Delta\text{CI}_{q-4\to q}$ (pp) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $\text{I(Miss CI)}_{q-4}$ | 4.639*** | 6.151*** | −0.788 | −0.744 | −0.564 | −0.567 | −0.217 |
| | (0.557) | (0.732) | (0.507) | (0.539) | (0.596) | (0.634) | (0.573) |
| $\text{CI}_{q-4}$ | | | −0.364*** | | | | |
| | | | (0.025) | | | | |
| Survey date FE | Yes | Yes | Yes | No | No | No | No |
| Forecaster FE | No | Yes | No | No | No | No | No |
| Survey date FE × CI | No | No | No | Exact | 1pp bins | 2pp bins | 5pp bins |
| Observations | 5,828 | 5,196 | 5,825 | 5,466 | 5,466 | 5,482 | 5,699 |
| Adjusted $R^2$ | 0.074 | 0.054 | 0.211 | 0.192 | 0.193 | 0.191 | 0.203 |

prediction can be expressed as follows:

$$\Delta\text{CI}_{i,q-4\to q} = \alpha + \beta \cdot \text{I(Miss CI)}_{i,q-4} + \epsilon_{i,q-4\to q}, \tag{1}$$

where $\Delta\text{CI}_{i,q-4\to q} = \text{CI}_{i,q} - \text{CI}_{i,q-4}$ is the change in the current forecast's CI relative to the same CFO's forecast four quarters ago, and $\text{I(Miss CI)}_{i,q-4}$ indicates whether the realized return missed the CI provided by the forecaster four quarters ago.

The results, reported in Table 5, initially appear consistent with our Bayesian learning model. Column (1) shows a positive coefficient on $\text{I(Miss CI)}_{q-4}$ that is statistically and economically significant. At face value, the results in Columns (1)–(2) indicate that CFOs learn quickly, adjusting their CIs upward or downward by 4.6pp–6.1pp with every hit or miss.

However, these results may reflect a spurious correlation due to endogeneity. The reason is that $\Delta\text{CI}_{q-4\to q}$ is mechanically related to $\text{I(Miss CI)}_{q-4}$. When $\text{CI}_{q-4}$ (which is part of $\Delta\text{CI}_{q-4\to q}$) is low due to reasons unrelated to fundamentals (e.g., measurement error),

$\text{I(Miss CI)}_{q-4}$ is likely to be high, and vice versa.

In other words, suppose a CFO randomly generates their confidence interval. With a very narrow CI draw, they will likely miss. The next random CI draw is mechanically more likely to be closer to the mean and, hence, wider. By construction, this is not learning, but this will look like learning in the baseline regression that does not control for endogeneity. We address this potential endogeneity by fully isolating the variation in $\text{I(Miss CI)}_{q-4}$ that is orthogonal to $\text{CI}_{q-4}$. One way to do so is to directly introduce $\text{CI}_{q-4}$ as a control variable, as done in Column (3). In this specification, the coefficient estimate on $\text{I(Miss CI)}_{q-4}$ is negative but becomes statistically and economically insignificant.

Another solution to this endogeneity problem is to compare forecasts of different CFOs with the same, albeit nonoverlapping, CIs. For example, consider two CIs provided on the same survey date and that have the same width of 10pp. One CI is $[-5\text{pp}, +5\text{pp}]$ and the other is $[-2\text{pp}, +8\text{pp}]$. Suppose that the realized return was $-3\%$. The first forecaster "hit" the CI, while the second forecaster "missed" it. Our model predicts that forecasters should learn from past performance and update their CIs accordingly. In the current example, forecasters' miss or hit outcomes differ, and thus they should update their CIs differently.

We implement this test in the regression by binning responses according to the survey date and initial CI. Consider first a simple case of one survey date. An exact approach would introduce indicators for each CI width (1pp, 2pp, . . . ) into the regression. Given multiple forecasts with the same CI width, the regression coefficient would capture the *within*-CI effect. In the "exact" specification that we estimate in Table 5, we include one indicator variable for each CI width that appears in the data. In other specifications, we loosen the grouping by allowing for binning: indicators for ranges of CIs of similar but not necessarily exact size; e.g., 2pp bins will put forecasts in bins based on their confidence intervals: 0pp–2pp, 2pp–4pp, etc. Given that we perform a panel regression with many survey dates, we need a set of CI indicators for each survey date. We, therefore, include interactions of survey fixed effects with CI indicators.

The endogeneity problem is resolved with the CI indicators. Because all forecasts in each bin share the same date and CI but have a potentially different outcome ($\text{I(Miss CI)}_{q-4}$), the dependent variable $\Delta \text{CI}_{q-4 \to q}$ is no longer mechanically related to the explanatory variable $\text{I(Miss CI)}_{q-4}$.

To ensure that our proposed approach is robust, we generate two simulated datasets: (*i*) simulated forecasters who follow the Bayesian model and learn from their earlier own performance, and (*ii*) forecasters who ignore their own past performance. Our tests discriminate well between the two simulations of updated CIs. Details on the simulations and the results of these tests are provided in Appendix D.

In Table 5, Columns (4) to (7), we present the results of regressions with CI–survey date bins. In Column (4), indicators are defined for each exact CI width. In Columns (5) to (7), indicators are defined based on bin sizes 1pp, 2pp, and 5pp wide, respectively. While retaining the negative coefficients, these estimates lose statistical and economic significance, failing to provide evidence consistent with the learning hypothesis.[13]

# 4 What Explains Miscalibration Levels?

So far, our study has not found evidence for improvement in calibration following repetition of the forecasting task or with past performance. Rather, our results show that miscalibration is largely an invariant personal trait that varies across CFOs. Note that "persistence" did not have to be the case for miscalibration; for example, miscalibration could have been random within-CFO, but we observe that some CFOs have varying degrees of miscalibration that persist in their forecasting.

In this section, we explore the cross-section of CFOs and test whether there is systematic

[13]In Appendix E, we provide additional specifications in which we control for the possibility that CFOs with the same CI but with different outlooks (as expressed by their expected returns) behave differently. In this analysis, we limit the sample to forecasts close to the median expected return in the respective survey. For example, in different specifications, we require expected returns to be within 1% or 5% of the median expected return. This analysis also reduces the chance that extreme observations drive our results. The results remain unchanged—none of the coefficients is statistically significant.

variation in measured CIs across several CFO characteristics: age, birth cohort, firm size (measured by revenue and employees), industry, and stock market expertise.

Here, we focus on the factors that may explain miscalibration measured as CIs rather than the hit rate. The reason is that our earlier results show that person-fixed effects best explain CIs. In contrast, the hit rate is best explained by unexpected stock market moves, captured by time-fixed effects (see Section 3.1). Our dependent variable in all the analyses in this section is adjusted-CI, i.e., adjusted for the average CI in that survey. This choice allows us to interpret the economic importance of different factors in explaining miscalibration.

## 4.1 Age and Birth Cohort

One possibility is that CFO age and birth cohort could explain the degree of observed miscalibration. Such association would be consistent with the literature proposing that a person's miscalibration tendency is driven by genetics and early life experiences and, hence is not easily adjusted by repetition or learning (Cesarini et al., 2009; Johnson and Fowler, 2011; Malmendier and Nagel, 2011; Malmendier et al., 2011).

In Figure 5, Panel (a), we plot the average CI (adjusted to survey-specific variations) as a function of CFO age. The figure shows that CIs decline with CFO age. CFOs aged 35 to 40 years old provide CIs with a width of 17%–18%. Those aged 40 to 60 provide CIs with a width of around 16%. Those in their 60s and 70s provide CIs with a width of around 13%–15%.

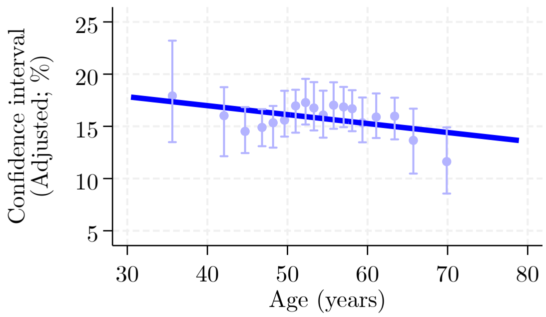
Since many CFOs complete multiple surveys over long periods, we can test whether the age effect that we observe reflects age or cohort effects, where "age effects" refer to individuals changing their CIs as they age or reflects "cohort effects" for individuals born in a certain year behaving differently than other birth-year cohorts. In Figure 5, Panel (b), we present the average CIs per birth-year cohort and note the cross-sectional variation.

In the first column of Table 6, we estimate that the explanatory power of CFO fixed effects in the subsample that includes age information is 0.606, implying that CFO characteristics

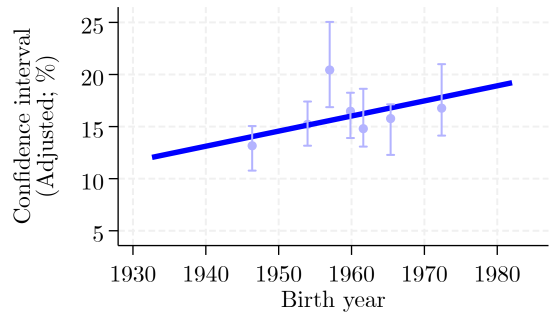
**Figure 5. Miscalibration by CFO Age and Birth Year**

This figure presents the relation between the average CI (percentage points, adjusted for survey fixed effects) with respect to a CFO's age (Panel (a)) and a CFO's birth year (Panel (b)). CFOs report their age in surveys conducted in 2008Q1 and 2014Q4, and their responses are propagated to other surveys. The generalized binscatter plots are created using the methodology and programs described in Cattaneo et al. (2024). $y$-axis variables are adjusted for calendar-time fixed effects (the sample average in each survey was added back).

(a) CFO age




(b) CFO birth year




explain a significant fraction of the cross-sectional variation in miscalibration. We estimate several additional specifications to better understand which variable—age or birth year—explains miscalibration levels. Age is not a statistically significant predictor of miscalibration (Columns (2) and (3)). Birth year is marginally statistically significant (Column (4)), and our estimates imply that CFOs with more recent birth years are better calibrated.

In a horserace that includes both age and birth year (Column (5)), the estimated impact of age remains statistically insignificant. In contrast, the estimated impact of birth year increases in magnitude and remains marginally significant. The estimates imply that for two CFOs of the same age but born 10 years apart (e.g., a 50-year-old born in 1970 vs. a 50-year-old born in 1960), the one born 10 years later (e.g., in 1970) has a confidence interval that is 2.6pp wider than the one born earlier, which is roughly 20% of the average confidence interval in the sample. This is economically meaningful but still nowhere near enough to obtain proper calibration.

In Appendix F, we explore life experiences as a mechanism to explain this cohort effect. Following the method in Malmendier and Nagel (2011), we regress each CFO's level of miscalibration on average lifetime-experienced realized volatility and average lifetime-experienced

**Table 6. Miscalibration and Personal Characteristics**

This table shows estimates from ordinary least squares regressions of CI (adjusted for survey average) on age and birth year. Standard errors are double-clustered by the survey's calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | CI (Adjusted; %) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Age | | −0.096 | | | 0.125 |
| | | (0.067) | | | (0.106) |
| log(Age) | | | −4.747 | | |
| | | | (3.629) | | |
| Birth year | | | | 0.145* | 0.260* |
| | | | | (0.077) | (0.133) |
| Forecaster FE | Yes | No | No | No | No |
| Observations | 4,850 | 4,963 | 4,963 | 4,963 | 4,963 |
| Adjusted $R^2$ | 0.606 | 0.005 | 0.004 | 0.010 | 0.012 |

Economic Policy Uncertainty (Bloom et al., 2017). While the results are statistically insignificant, they are consistent with the idea that early life experiences with heightened aggregate uncertainty, captured by birth cohort, may have lasting impacts on miscalibration.

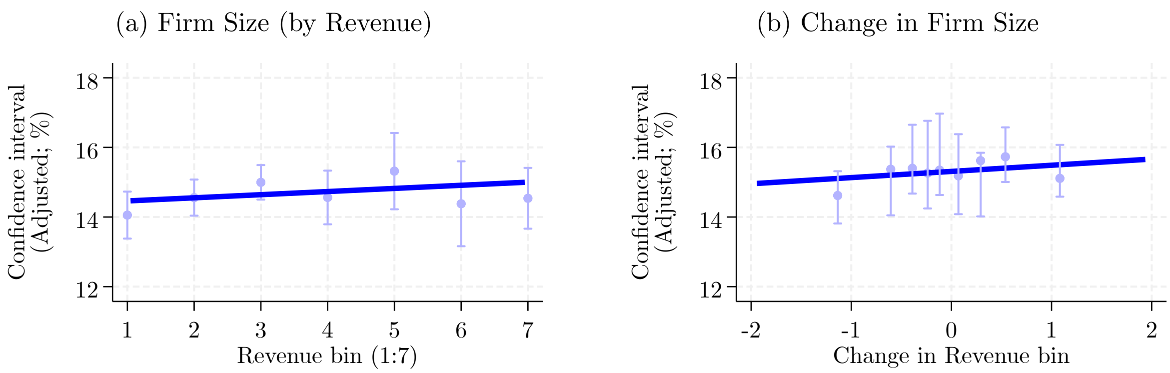
## 4.2 Firm Size

We next examine the relation between miscalibration and firm size. We plot the average CFO's adjusted CI as a function of the firm revenue bin (ranging from 1 to 7; see Section 1.3). Figure 6, Panel (a), shows that CFOs who work in large firms show better calibration than their peers in smaller firms. However, while the effect is statistically significant, its economic magnitude is moderate. The average CI at the smallest firms is about 12.2%, and increases to only about 15% for CFOs at the largest firms.

The negative correlation between firm size and miscalibration does not necessarily indicate causality. One possibility is that large firms will employ individuals who demonstrate better calibration. In contrast, CFOs with more severe miscalibration may prefer to be employed by smaller/start-up firms, akin to self-sorting based on psychological biases (Gervais

**Figure 6. CFO Miscalibration and Firm Size**

This figure presents the relation between the average CI (in percentage points, adjusted for survey fixed effects) with respect to firm size bins. CFOs are asked in each survey about their firm's revenues, on a scale of 1–7: (1) <\$25m, (2) \$25m–\$100m, (3) \$100m–\$500m, (4) \$500m–\$1bn, (5) \$1bn–\$5bn, (6) \$5bn–\$10bn, and (7) >\$10bn. The generalized binscatter plots are created using the methodology and programs described in Cattaneo et al. (2024). $y$-axis variables are adjusted for calendar-time fixed effects (the sample average was added back).




and Goldstein, 2007; Puri and Robinson, 2007; Levy and Tasoff, 2017).

Our survey can provide some insight into the mechanism that drives this correlation. Because CFOs respond to our surveys multiple times over long periods, we can track the evolution of their CIs as the size of their firm evolves (e.g., due to changing jobs to smaller or larger firms).

In Figure 6, Panel (b), we limit the sample to forecasts made by CFOs who reported different firm sizes over time. We then plot their CIs as a function of the deviation from their own average firm size. Naturally, most CFOs move up or down one or two notches in the firm-size scale. Overall, within a CFO's career, we do not observe a material difference in CIs across firm sizes.

We obtain similar results in regression form. Table 7, Column (2) shows that CFOs' CIs increase with firm size. Based on the coefficient, a shift from a small firm (bin 1) to a very large firm (bin 7) is associated with an increase in CI width of $0.465 \times (7 - 1) = 2.8$pp. Column (3) shows that CFO-fixed effects absorb the entire effect. After CFO-fixed effects are included, the remaining effect is $0.09 \times (7 - 1) = 0.6$pp and statistically insignificant.

**Table 7. Miscalibration and Corporate Environment**

This table shows estimates from ordinary least squares regressions of CI (adjusted for survey average) on firm size and industry. Standard errors are double-clustered by the survey's calendar date and participant. In Columns (2), (3), (5), and (6), the omitted category is size bin 1 (smallest firms). In Columns (7) and (8), the omitted category is Retail/Wholesale. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | CI (Adjusted; %) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Size (revenue; 1–7) | | 0.465*** | 0.092 | | | | | |
| | | (0.099) | (0.091) | | | | | |
| Size (employees; 1–7) | | | | | 0.216** | 0.191 | | |
| | | | | | (0.090) | (0.164) | | |
| Mining/Construction | | | | | | | 0.203 | 0.320 |
| | | | | | | | (1.040) | (0.851) |
| Manufacturing | | | | | | | 0.650 | 0.508 |
| | | | | | | | (0.641) | (0.702) |
| Transportation/Energy | | | | | | | 2.958** | 0.158 |
| | | | | | | | (1.216) | (0.838) |
| Communications/Media | | | | | | | −1.102 | 0.952 |
| | | | | | | | (0.967) | (0.988) |
| Tech (Software/Biotech) | | | | | | | 0.511 | 0.038 |
| | | | | | | | (0.851) | (0.777) |
| Banking/Finance/Insurance | | | | | | | 2.743*** | −0.850 |
| | | | | | | | (0.838) | (0.917) |
| Service/Consulting | | | | | | | 0.247 | 0.163 |
| | | | | | | | (0.717) | (0.633) |
| Healthcare/Pharmaceutical | | | | | | | −0.203 | 1.460 |
| | | | | | | | (0.800) | (1.046) |
| Other | | | | | | | 0.215 | 0.160 |
| | | | | | | | (0.686) | (0.507) |
| Forecaster FE | Yes | No | Yes | Yes | No | Yes | No | Yes |
| Observations | 17,053 | 20,662 | 17,053 | 12,091 | 15,362 | 12,091 | 20,700 | 17,056 |
| Adjusted $R^2$ | 0.583 | 0.004 | 0.583 | 0.578 | 0.001 | 0.578 | 0.008 | 0.582 |

The effects of firm size on the adjusted-CI appear more likely to be due to the sorting of CFOs into firms rather than a causal influence of firm environment on CFO miscalibration. We see that when we include firm size indicators in a regression with CFO fixed effects, these size indicators do not meaningfully change the $R^2$ obtained from CFO fixed effects alone (compare Columns (1) and (3)). This implies that CFOs moving across firm sizes do not change their adjusted CI. Hence, the initial impact reported in Column (2) is likely to be driven by the sorting of better-calibrated CFOs into larger organizations.

We repeat this procedure for firm size measured by the number of employees. The results in Columns (4) to (6) are qualitatively similar.

## 4.3 Industry

We also have a broad industry classification for CFOs. As with firm size, we notice a small effect of industries on the adjusted CI; there are two industries with significant effects and $R^2$ of 0.008 in Column (7). As in the case of firm size, industry effects seem more likely to be driven by CFOs sorting into specific industries. When including CFO fixed effects in Column (8), industry effects become statistically insignificant, and they have no added explanatory power to the $R^2$ of CFO fixed effects alone.

Taken together, our results provide no evidence that CFOs' corporate environments appear to shape their miscalibration. At the same time, we note that an important caveat to drawing strong conclusions from this analysis is that CFO movement across firm size is endogenous.

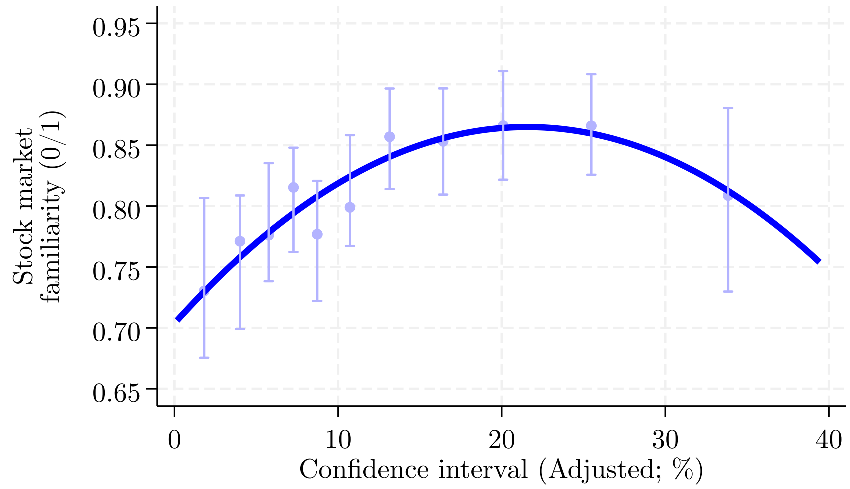
## 4.4 Stock Market Familiarity

The final characteristic we study is CFOs' self-reported familiarity with the stock market. As discussed in Section 1.3, since 2020, we have asked CFOs to assess their own familiarity with the stock market. In Figure 7, we calculate the fraction of CFOs who have high familiarity with the stock market across the CI spectrum.

CFOs who are more miscalibrated or overconfident with respect to stock market returns (i.e., they have narrower confidence intervals) have the lowest stock market familiarity. This finding is reminiscent of the seminal Dunning-Kruger effect that less competent survey respondents are overconfident in their own abilities (Kruger and Dunning, 1999): respondents who should perhaps be the most conservative with their responses are instead the most assertive. In our context, the measure of "competence" is stock market familiarity, and the measure of "confidence" is the degree of miscalibration; CFOs, with the least familiarity,

**Figure 7. Miscalibration and S&P Familiarity**

This figure presents the relation between the average S&P 500 familiarity variable with respect to the reported CI (expressed in percentage points). This generalized binscatter plot was created using the methodology and programs described in Cattaneo et al. (2024). CIs ($x$-axis variable) are adjusted for calendar time fixed effects (the sample average was added back). The fitted curve is a polynomial of second degree.




provide the tightest confidence intervals.

We also note that the data suggest an inverse U-shape: the most well-calibrated CFOs are also less likely to be familiar with the S&P 500.[14] The inverse U-shaped relation between stock market familiarity and miscalibration could also reflect a trade-off between two opposing forces. In one direction, those familiar with the stock market may be more acutely aware of the market's volatility and, as a result, show less miscalibration. On the other hand, increased familiarity may lead to a false sense of confidence and, in turn, to narrower CIs. The U-shaped pattern is consistent with the former force dominating for moderate familiarity and the latter for high levels of familiarity.

[14]The quadratic polynomial fits the data better than a linear polynomial. To ensure this is not driven by extreme observations, we confirm the inverse U-shape even when restricting the sample to exclude CIs wider than 20pp, 25pp, 30pp, or 35pp.

# 5 Conclusion

The beliefs of senior financial executives about the mean of the distribution of S&P 500 returns are unbiased, but they greatly underestimate the variance of this distribution. This miscalibration is a type of overconfidence and can be measured in our sample as the confidence intervals that CFOs provide around their point estimate forecasts or the ex-post hit rate of realized returns onto their confidence intervals.

Our analysis shows that CFOs' miscalibration is highly persistent along multiple dimensions. At the aggregate level, we do not find evidence that calibration has improved with the greater general awareness of behavioral biases over the last two decades. At the individual level, we do not find evidence that forecasters show improved calibration as they repeat their forecasting tasks. We also find no evidence consistent with rational learning about stock market volatility from past stock market realizations according to Bayes' rule. Instead, miscalibration is highly persistent and, therefore, appears to be a CFO-level trait.

We explore factors that may explain the persistence of miscalibration. CFOs' miscalibration levels are associated with their birth cohorts and we analyze lifetime experiences as a potential source of miscalibration. We document a relation between CFO miscalibration and familiarity with the stock market that is reminiscent of the Dunning-Kruger effect: compared to the middle of the miscalibration distribution, overconfident CFOs with high miscalibration are typically the least familiar with the stock market, while CFOs with (relatively) low miscalibration are also less familiar with the stock market.

Our findings contribute to the broader literature on miscalibration, overconfidence, and learning in decision-making. Our results suggest that miscalibration persists even in a professional setting where individuals can learn from past experiences. This has important implications for understanding the limits of learning in mitigating cognitive biases and designing policies and interventions to address such biases.

Furthermore, our results have practical implications for financial markets and corporate decision-making. Miscalibration among top executives has first-order impacts on investment and financing decisions, and its persistence may affect firm value in the long run. Recognizing the limitations of feedback in reducing miscalibration could prompt organizations to adopt alternative approaches, such as employing decision-support tools that counteract biases or explicitly accounting for such biases in decision-making processes.

# References


Alpert, Marc, and Howard Raiffa, 1982, A progress report on the training of probability assessors, in Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases*, 294–305 (Cambridge University Press).

Baranski, Joseph V., and William M. Petrusic, 1999, Realism of confidence in sensory discrimination, *Perception & Psychophysics* 61, 1369–1383.

Barrero, Jose Maria, 2022, The micro and macro of managerial beliefs, *Journal of Financial Economics* 143, 640–667.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey, 2013, Managerial miscalibration, *Quarterly Journal of Economics* 128, 1547–1584.

Benson, P. George, and Dilek Önkal, 1993, The effects of feedback and training on the performance of probability forecasters, *International Journal of Forecasting* 8, 559–573.

Bloom, Nicholas, Stephen J. Davis, Lucia Foster, Brian Lucking, Scott Ohlmacher, and Itay Saporta-Eksten, 2017, Business-level expectations and uncertainty, Working paper, National Bureau of Economic Research.

Campbell, Sandy, and Don A. Moore, 2024, Overprecision in the Survey of Professional Forecasters, *Collabra: Psychology* 10.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, 2024, On binscatter, *American Economic Review* 114, 1488—-1514.

Cesarini, David, Paul Lichtenstein, Magnus Johannesson, and Björn Wallace, 2009, Heritability of overconfidence, *Journal of the European Economic Association* 7, 617–627.

Chiang, Yao-Min, David Hirshleifer, Yiming Qian, and Ann E. Sherman, 2011, Do investors learn from experience? Evidence from frequent IPO investors, *Review of Financial Studies* 24, 1560–1589.

Deshmukh, Sanjay, Anand M. Goel, and Keith M. Howe, 2013, CEO overconfidence and dividend policy, *Journal of Financial Intermediation* 22, 440–463.

Feng, Lei, and Mark S. Seasholes, 2005, Do investor sophistication and trading experience eliminate behavioral biases in financial markets?, *Review of Finance* 9, 305–351.

Fermand, Elyas, Camelia M. Kuhnen, Geng Li, and Itzhak Ben-David, 2024, Extrapolative uncertainty and household economic behavior, *Management Science* 70, 5607–5625.

Galasso, Alberto, and Timothy S. Simcoe, 2011, CEO overconfidence and innovation, *Management Science* 57, 1469–1484.

Gennaioli, Nicole, Yueran Ma, and Andrei Shleifer, 2016, Expectations and investment, in Martin Eichenbaum, and Jonathan Parker, eds., *NBER Macroeconomics Annual*, volume 30, 379–431 (National Bureau of Economic Research).

Gervais, Simon, and Itay Goldstein, 2007, The positive effects of biased self-perceptions in firms, *Review of Finance* 11, 453–496.

Gervais, Simon, and Terrance Odean, 2001, Learning to be overconfident, *Review of Financial Studies* 14, 1–27.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five facts about beliefs and portfolios, *American Economic Review* 111, 1481–1522.

Goel, Anand M., and Anjan V. Thakor, 2008, Overconfidence, CEO selection, and corporate governance, *Journal of Finance* 63, 2737–2784.

González-Vallejo, Claudia, and Aaron Bonham, 2007, Aligning confidence with accuracy: Revisiting the role of feedback, *Acta Psychologica* 125, 221–239.

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

Graham, John R., Campbell R. Harvey, and Manju Puri, 2013, Managerial attitudes and corporate actions, *Journal of Financial Economics* 109, 103–121.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Hirshleifer, David, Angie Low, and Siew Hong Teoh, 2012, Are overconfident CEOs better innovators?, *Journal of Finance* 67, 1457–1498.

Johnson, Dominic D.P., and James H. Fowler, 2011, The evolution of overconfidence, *Nature* 477, 317–320.

Kaustia, Markku, and Samuli Knüpfer, 2008, Do investors overweight personal experience? Evidence from IPO subscriptions, *Journal of Finance* 63, 2679–2702.

Keefer, Donald L., and Samuel E. Bodily, 1983, Three-point approximations for continuous random variables, *Management Science* 29, 595–609.

Kruger, Justin, and David Dunning, 1999, Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments, *Journal of Personality and Social Psychology* 77, 1121–1134.

Kuchler, Theresa, and Basit Zafar, 2019, Personal experiences and expectations about aggregate outcomes, *Journal of Finance* 74, 2491–2542.

Kuhnen, Camelia M., 2014, Asymmetric learning from financial information, *Journal of Finance* 70, 2029–2062.

Levy, Matthew R., and Joshua Tasoff, 2017, Exponential-growth bias and overconfidence, *Journal of Economic Psychology* 58, 1–14.

Lichtenstein, Sarah, and Baruch Fischhoff, 1980, Training for calibration, *Organizational Behavior and Human Performance* 26, 149–171.

List, John A., 2007, Field experiments: A bridge between lab and naturally occurring data, *B.E. Journal of Economic Analysis & Policy* 6, 0000102202153806371747.

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

Malmendier, Ulrike, and Geoffrey Tate, 2005, CEO overconfidence and corporate investment, *Journal of Finance* 60, 2661–2700.

Malmendier, Ulrike, and Geoffrey Tate, 2008, Who makes acquisitions? CEO overconfidence and the market's reaction, *Journal of Financial Economics* 89, 20–43.

Malmendier, Ulrike, and Geoffrey Tate, 2015, Behavioral CEOs: The role of managerial overconfidence, *Journal of Economic Perspectives* 29, 37–60.

Malmendier, Ulrike, Geoffrey Tate, and Jon Yan, 2011, Overconfidence and early life experiences: The effect of managerial traits on corporate financial policies, *Journal of Finance* 66, 1687–1733.

Mannes, Albert E., and Don A. Moore, 2013, A behavioral demonstration of overconfidence in judgment, *Psychological Science* 24, 1190–1197.

Martin, Ian W.R., and Dimitris Papadimitriou, 2022, Sentiment and speculation in a market with heterogeneous beliefs, *American Economic Review* 112, 2465–2517.

McKenzie, Craig R.M., Michael J. Liersch, and Ilan Yaniv, 2008, Overconfidence in interval estimates: What does expertise buy you?, *Organizational Behavior and Human Decision Processes* 107, 179–191.

Meeuwis, Maarten, Jonathan .A Parker, Antoinette Schoar, and Duncan Simester, 2022, Belief disagreement and portfolio choice, *Journal of Finance* 77, 3191–3247.

Mood, Alexander McFarlane, Franklin A. Graybill, and Duane C. Boes, 1974, *Introduction to the Theory of Statistics*, third edition (McGraw-Hill).

Moore, Don A., and Amelia S. Dev, 2018, Individual differences in overconfidence, in V. Zeigler-Hill, and T. K. Shackelford, eds., *Encyclopedia of Personality and Individual Differences* (Springer).

Moore, Don A., and Paul J. Healy, 2008, The trouble with overconfidence, *Psychological Review* 115, 502.

Moore, Don A., and Derek Schatz, 2017, The three faces of overconfidence, *Social and Personality Psychology Compass* 11, e12331.

Moore, Don A., and Samuel A. Swift, 2011, The three faces of overconfidence in organizations, in D. De Cremer, R. van Dick, and J. K. Murnighan, eds., *Social Psychology and Organizations*, 147–184 (Routledge/Taylor & Francis Group).

Moore, Don A., Elizabeth R. Tenney, and Uriel Haran, 2015, Overprecision in judgment, in Gideon Keren, and George Wu, eds., *Handbook of Judgement and Decision Making* (Oxford: Wiley Blackwell).

Newey, Whitney K., and Kenneth D. West, 1986, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Working paper, National Bureau of Economic Research.

Pflueger, Carolin, Emil Siriwardane, and Adi Sunderam, 2020, Financial market risk perceptions and the macroeconomy, *Quarterly Journal of Economics* 135, 1443–1491.

Puri, Manju, and David T. Robinson, 2007, Optimism and economic choice, *Journal of financial economics* 86, 71–99.

Rakow, Tim, Nigel Harvey, and Sarah Finer, 2003, Improving calibration without training: The role of task information, *Applied Cognitive Psychology* 17, 419–441.

Sieck, Winston R., and Hal R. Arkes, 2005, The recalcitrance of overconfidence and its contribution to decision aid neglect, *Journal of Behavioral Decision Making* 18, 29–53.

Soll, Jack B., and Joshua Klayman, 2004, Overconfidence in interval estimates, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30, 299–314.

Soll, Jack B., Katherine L. Milkman, and John W. Payne, 2015a, Outsmart your own biases, *Harvard Business Review* 93, 64–71.

Soll, Jack B., Katherine L. Milkman, and John W. Payne, 2015b, A user's guide to debiasing, in *The Wiley Blackwell handbook of judgment and decision making*, volume 2, 924–951 (Wiley Online Library).

Subbotin, Vadim, 1996, Outcome feedback effects on under- and overconfident judgments (general knowledge tasks), *Organizational Behavior and Human Decision Processes* 66, 268–276.

Thaler, Richard H., 1987, The psychology of choice and the assumptions of economics, in Alvin E. Roth, ed., *Laboratory experimentation in economics: Six points of view*, 99–130 (Cambridge University Press Cambridge).

Thaler, Richard H., 1994, *Quasi Rational Economics* (Russell Sage Foundation).

Thaler, Richard H., 2000, From homo economicus to homo sapiens, *Journal of Economic Perspectives* 14, 133–141.

# Appendix A Sample Demographics

In Table A.1, we summarize demographics from our survey, when available to those in Compustat and Execucomp. We draw data from these sources for 2001 to 2022, corresponding to our survey sample. For Compustat, we require firms to be publicly traded on the NYSE, NASDAQ, or AmEx. We benchmark to industry distribution, sales, and the number of employees. Industries are linked based on the Fama-French 49-industry classification (FF49 in the table). Execucomp covers firms in the S&P 1500 index. Observations are limited to those identified as CFOs and are used to determine the age distribution of CFOs.

## Table A.1. Benchmarking Sample Demographics to U.S. Public Firms

The table presents the distribution of demographic information for survey respondents (at the individual forecast level) compared to the distribution of similar variables for the universe of U.S. publicly traded firms. The data come from Compustat and Execucomp. Compustat is used to benchmark industry distribution, sales, and the number of employees. Execucomp is used to determine the age distribution of CFOs.

| | | Sample | | | |
|---|---|---|---|---|---|
| | | All | Public | Non-public | Compustst‌ |
| Industry: | Retail/Wholesale (FF49: 42, 43) | 12.4% | 7.3% | 13.9% | 7.0% |
| | Mining/Construction (FF49: 18, 28–30) | 5.1% | 3.7% | 5.3% | 4.8% |
| | Manufacturing (FF49: 1–10, 13–27, 39, 40) | 23.6% | 27.8% | 23.0% | 30.2% |
| | Transportation/Energy (FF49: 31, 41) | 5.2% | 8.2% | 4.6% | 4.4% |
| | Communications/Media (FF49: 7, 32) | 3.0% | 5.2% | 2.7% | 2.4% |
| | Tech (Software/Biotech) (FF49: 35–38) | 5.5% | 7.7% | 5.3% | 17.9% |
| | Banking/Finance/Insurance (FF49: 45–48) | 13.6% | 19.6% | 11.7% | 19.8% |
| | Service/Consulting (FF49: 33, 34, 47, 49) | 13.7% | 5.7% | 12.8% | 6.6% |
| | Healthcare/Pharmaceutical (FF49: 11, 12) | 6.1% | 6.7% | 6.6% | 5.3% |
| | Other (FF49: 44) | 11.7% | 8.2% | 14.3% | 1.8% |
| | Observations | 27,806 | 5,872 | 17,139 | 80,761 |
| Sales: | <$25m | 21.6% | 3.2% | 24.0% | 12.5% |
| | $25m–$100m | 24.0% | 8.8% | 29.3% | 21.0% |
| | $100m–$500m | 26.0% | 19.1% | 30.2% | 24.0% |
| | $500m–$1bn | 8.0% | 12.7% | 7.3% | 10.8% |
| | $1bn–$5bn | 11.2% | 27.2% | 6.8% | 19.2% |
| | $5bn–$10bn | 4.0% | 12.9% | 1.3% | 5.1% |
| | >$10bn | 5.3% | 16.2% | 1.1% | 7.5% |
| | Observations | 27,949 | 5,823 | 17,289 | 117,147 |
| Employees: | <100 | 21.8% | 3.8% | 28.0% | 15.7% |
| | 100–499 | 29.9% | 14.2% | 35.5% | 22.7% |
| | 500–999 | 11.9% | 7.9% | 13.4% | 10.7% |
| | 1000–2499 | 11.2% | 11.8% | 11.1% | 13.6% |
| | 2500–4999 | 7.2% | 12.2% | 5.4% | 10.3% |
| | 5000–9999 | 5.6% | 13.7% | 2.7% | 8.9% |
| | >10,000 | 12.4% | 36.4% | 3.9% | 18.0% |
| | Observations | 21,929 | 5,530 | 15,634 | 105,642 |
| Age: | 25–34 | 2.0% | 1.6% | 2.1% | 0.5% |
| | 35–44 | 12.5% | 21.8% | 11.5% | 16.2% |
| | 45–54 | 40.8% | 45.7% | 40.6% | 50.8% |
| | 55–64 | 34.4% | 23.9% | 35.2% | 30.2% |
| | 65–74 | 9.6% | 6.9% | 10.0% | 2.2% |
| | 75–84 | 0.4% | 0.0% | 0.4% | 0.1% |
| | 85+ | 0.2% | 0.0% | 0.3% | 0.0% |
| | Observations | 5,004 | 669 | 3,799 | 33,792 |

# Appendix B Bayesian Learning: Framework

In Section 2, we present a framework that describes how beliefs over the return distribution are formed from primitive beliefs over the average return and the standard deviation of returns. In what follows, we parameterize the framework using standard assumptions in the Bayesian learning literature. The parameterized model generates testable predictions that we take to the data.

## B.1 How Do Bayesian Forecasters Form Their Beliefs?

Figure B.1 presents the belief formation framework for forecasters that believe returns are normally distributed and thus form beliefs over the two parameters that fully characterize the distribution: the mean, $\bar{r}$, and variance, $\sigma_r^2$. In the figure, the black captions describe the general framework, and the blue captions detail a standard Bayesian parameterization of the process framework.

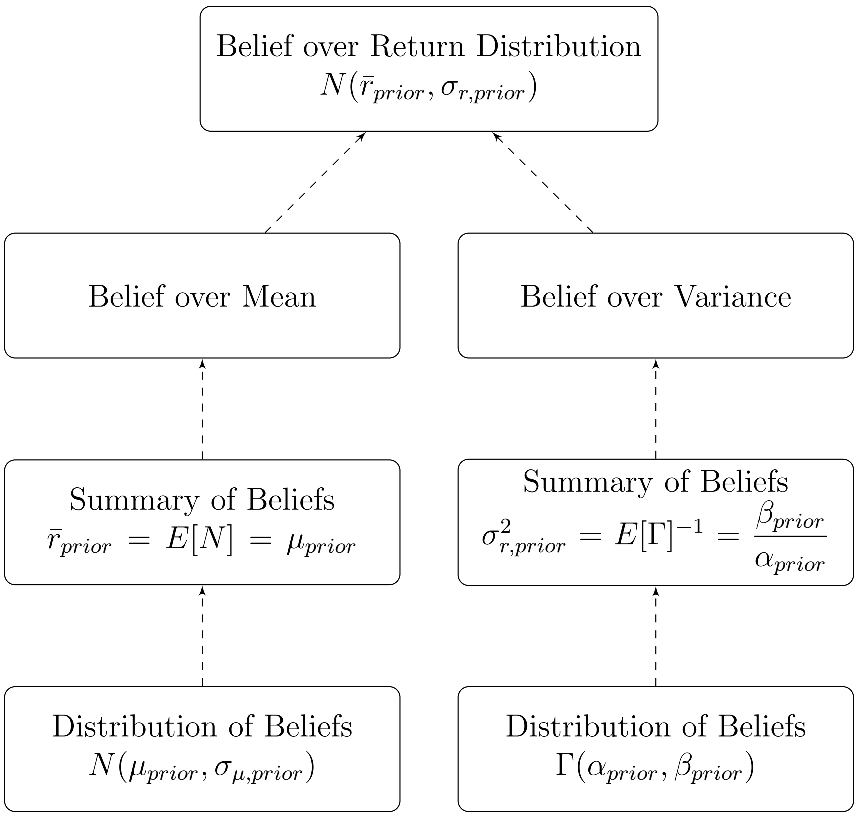
Our main survey question elicits a CFO's belief of the first two moments of the return distribution: the mean and the variance. Since CFOs are not certain about these moments, we assume that the beliefs of these quantities are also distributions. These summary beliefs that CFOs report are each derived from two underlying distributions: one for the mean and the other for the variance. To maintain clarity, we will refer to the mean of each belief distribution as the "belief," and the variance of each belief distribution as the "strength," "tightness," or "conviction" with which the belief is held. Thus, we are interested in the sources, or priors, of four distinct objects: the belief of the mean, the strength of the belief of the mean, the belief of the variance, and the strength of the belief of the variance.[15]

Table B.1 summarizes the belief distributions and their sources. In our framework, CFOs have prior beliefs over two distinct parameters, namely, the mean of the return distribution and the variance of the return distribution. In turn, a distribution of beliefs is formed for

[15]Respectively, these are technically the mean and variance of the distribution of beliefs of the mean and the mean and variance of the distribution of beliefs of the variance.

**Figure B.1. Framework with Paramaterized Distributions: Formation of Beliefs over Return Distribution**

This figure illustrates the relation between the belief distributions over the mean and variance and the belief over the return distribution. Beginning from the lower left panel, the unknown belief over the mean is a normally distributed random variable. Moving up one panel, this distribution is summarized by its first moment, which forms the belief over the mean. In the lower right panel, the unknown belief over the variance is a gamma-distributed random variable, again summarized by its first moment. Together, these two beliefs form a belief over the return distribution, summarized in the topmost panel.




each parameter, and each of these distributions is characterized by a belief (i.e., mean) and tightness (i.e., variance).

Taken together, this framework allows the survey responses to be mapped directly to CFOs' beliefs regarding mean and variance.

## B.2 Prior Beliefs of Mean and Variance

Focusing first on the left side of Figure B.1, the survey question regarding the point forecast elicits the belief over the mean, $\bar{r}_{prior}$. Similarly, on the right side of the figure, the questions about the confidence interval elicit the belief over the standard deviation, $\sigma_{r,prior}$.

**Table B.1. Summary of Belief Distributions and Sources of Beliefs**

| Unknown parameter | Moment of belief distribution | Label | Plausible source of belief |
|---|---|---|---|
| Mean | Mean | Belief over mean | Sample mean |
| | Variance | Conviction of belief | Variance of sample mean |
| Variance | Mean | Belief over variance | Sample variance |
| | Variance | Conviction of belief | Variance of sample variance |

These are the first moments of each belief distribution and are sufficient to generate forecasts of the future S&P 500 return distribution. The natural candidate for the belief over the mean is the sample mean of the historical returns. Similarly, the sample variance is the natural candidate for the belief over the variance.

## B.3 Conviction of Prior Beliefs

We now turn to the tightness or strength of conviction of beliefs. These are the second moments of the distributions of beliefs over the mean and variance of the return distribution. What are the natural candidates for the tightness of the belief distributions? Recall that the mean and variance of beliefs are derived from the sample mean and variance (of the historical return distribution). Since these sample moments are random variables, their variances are natural candidates for tightness. In particular, the belief over the mean is given by the sample mean, and thus the tightness is given by the variance of the sample mean, $\frac{\sigma_r^2}{N_r}$, where $\sigma_r^2$ is the (unknown) population variance. Similarly, the belief over the variance is given by the variance of the sample. Thus, the tightness is given by the variance of the sample skewness, $\frac{2\sigma_r^4}{N_r-1}$.[16]

Since the population variance is unknown, it is estimated using the sample variance. Both of these variances increase in the sample variance. Taken together, the sample variance is used to inform the prior beliefs in three ways: the belief over the variance and the tightness

[16] See Mood, Graybill, and Boes (1974) for a detailed derivation of the general case (p. 229). The case in which returns are normally distributed is presented here for ease of exposition. In the more general case, the variance of the sample variance is given by $\left[\frac{\kappa}{N_r} + \frac{2}{N_r-1}\right]\sigma^4$, where $\kappa$ is defined as excess kurtosis.

of beliefs about both the mean and the variance.

The analysis above highlights that CFOs' beliefs about the sample variance are much smaller than indicated by the historical data. As a result, since the variances of the belief distributions are simply increasing functions of the belief over the sample variance, it follows that CFOs' variances of the beliefs over the mean and variance are also much smaller than the data suggest they should be. CFOs have a firm conviction in their beliefs, referred to as "tight priors" in the Bayesian literature. Note that miscalibration arises when the mean of the belief over the variance is smaller (or larger) than what the realized data would suggest, while a strong conviction arises when the variances of the beliefs over the mean and variance are smaller than what the historical data would suggest.

# Appendix C Selection Into and Out of the Survey

In our main analysis, we perform a matching exercise instead of pooling all responses. This is important to ensure that the empirical results are not driven by "composition effects" of changes in CI widths by two forecasters with vastly different initial CIs. In this appendix, we use the same methodology to test whether CFOs who miss are more likely to drop out of the sample. In Table C.1, we regress an indicator for whether the CFO exited from the survey, that is, whether the observed response was their last response, on whether they hit the interval.

**Table C.1. Sample Exit and Forecast Performance**

This table shows estimates from ordinary least squares regressions of an indicator for whether the observation is the CFO's last forecast on an indicator for whether the forecast hit the interval. Standard errors are double-clustered by the survey's calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent variables: | I(Last forecast)$_q$ ×100 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(Miss CI)$_{q-4}$ | 0.115 | 0.262 | 0.344 | 0.345 |
| | (1.826) | (1.976) | (1.966) | (1.729) |
| Survey date FE | No | No | No | No |
| Survey date × CI FE | Exact | 1pp bins | 2pp bins | 5pp bins |
| Observations | 4,801 | 4,801 | 4,814 | 5,019 |
| Adjusted $R^2$ | 0.020 | 0.019 | 0.018 | 0.020 |

If CFOs who miss are more likely to exit the sample, we should see a positive and significant coefficient. Alternatively, if CFOs who are hit are more likely to exit the sample, we should see a negative and significant coefficient. We find no relation between hitting the CI and exiting the sample. The coefficients are slightly negative but insignificant, consistent with the view that CFOs who show more miscalibration are (in general) more likely to drop out of the sample.

# Appendix D Matching Validation via Simulated Data

In this appendix, we test whether introducing fixed effects for each confidence interval width by survey date indicator pair can capture CFOs' learning. We do so by simulating the CIs that CFOs would have provided under two distinct scenarios.

Under the "random confidence interval" scenario, CFOs generate CIs that are not informed by their past hits/misses,

$$RandomCI_{i,q} = CI_{i,q-4} \cdot (0.7 + 0.3 \cdot (0.2 \cdot RAND[-1, +1])), \tag{2}$$

where $RAND[-1, +1]$ is a random variable that receives the value $-1$ or $+1$ with equal probability. Under this scenario, the current CI is a weighted average of the lagged CI and the zero mean noise. The change in the confidence interval is now defined as $\Delta CI^{Random}_{i,q-4\to q} = RandomCI_{i,q} - CI_{i,q-4}$.

Under the "learning-based confidence interval" scenario, CFOs generate CIs that are informed by their past hits/misses:

$$\begin{aligned} LearningCI_{i,q} \;\; = \;\; & CI_{i,q-4} \cdot [0.7 + 0.3 \cdot (0.2 \cdot RAND[-1, +1]) \\ & +0.1 \cdot (I(MissCI)_{i,q-4} - I(HitCI)_{i,q-4})], \end{aligned} \tag{3}$$

where $RAND[-1, +1]$ is defined as above and $I(MissCI)_{i,q-4}$ and $I(HitCI)_{i,q-4}$ are variables indicating whether the previous forecast as a miss or a hit. These variables induce feedback: following a past miss, the current CI widens, whereas following a hit, the current CI narrows. The change in the CI is defined now as $\Delta CI^{Learning}_{i,q-4\to q} = LearningCI_{i,q} - CI_{i,q-4}$.

In Table D.1, we repeat the test in Table 5 using the dependent variables $\Delta CI^{Random}_{i,q-4\to q}$ and $\Delta CI^{Learning}_{i,q-4\to q}$ in Panels A and B, respectively. The results show that the endogenous specification (Column (1)) detects statistically and economically significant evidence for learning in both panels, regardless of whether the dependent variable reflects true learning. In contrast, the specifications that control for the lagged CIs (Columns (3) to (7)) discriminate well between the dependent variables that are based on random confidence intervals, that is, no learning (Panel A) or realization-based learning (Panel B).

### Table D.1. Validation of Specifications to Test for Learning

This table shows estimates from ordinary least squares regressions of miscalibration measures on the forecast number. Standard errors are double-clustered by the survey's calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

**Panel A: Tests Using Simulated Data with No Learning**

| Dependent variable: | $\Delta CI^{Random}_{q-4\to q}$ (%; Simulated; no learning) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $I(\text{Miss CI})_{q-4}$ | 4.468*** | 1.744*** | 0.028 | 0.022 | 0.114 | 0.136* | 0.201** |
| | (0.255) | (0.196) | (0.051) | (0.050) | (0.072) | (0.074) | (0.088) |
| $CI_{q-4}$ | | | −0.298*** | | | | |
| | | | (0.002) | | | | |
| Survey date FE | Yes | Yes | Yes | No | No | No | No |
| Forecaster FE | No | Yes | No | No | No | No | No |
| Survey date FE × CI | No | No | No | Exact | 1pp bins | 2pp bins | 5pp bins |
| Observations | 5,825 | 5,196 | 5,825 | 5,466 | 5,466 | 5,482 | 5,699 |
| Adjusted $R^2$ | 0.212 | 0.586 | 0.904 | 0.908 | 0.890 | 0.891 | 0.871 |

**Panel B: Tests Using Simulated Data with Forced Learning**

| Dependent variable: | $\Delta CI^{Learning}_{q-4\to q}$ (%; Simulated; with learning) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $I(\text{Miss CI})_{q-4}$ | 8.225*** | 5.832*** | 3.942*** | 3.590*** | 3.608*** | 3.623*** | 3.919*** |
| | (0.500) | (0.522) | (0.472) | (0.414) | (0.447) | (0.448) | (0.458) |
| $CI_{q-4}$ | | | -0.287*** | | | | |
| | | | (0.020) | | | | |
| Survey date FE | Yes | Yes | Yes | No | No | No | No |
| Forecaster FE | No | Yes | No | No | No | No | No |
| Survey date FE × CI | No | No | No | Exact | 1pp bins | 2pp bins | 5pp bins |
| Observations | 5,825 | 5,196 | 5,825 | 5,467 | 5,467 | 5,481 | 5,699 |
| Adjusted $R^2$ | 0.181 | 0.271 | 0.348 | 0.374 | 0.374 | 0.376 | 0.385 |

In summary, the results here show that specifications that control for the lagged CI, either directly or through interactions, appear to resolve the endogeneity problem discussed in Section 3.3. In other words, these specifications provide solid empirical tests of whether forecasters in the dataset update their CIs based on past realizations.

# Appendix E Learning Hypothesis: Robustness Tests

In Appendix Table E.1, we test for the robustness of the regression specifications in Table 5, Columns (4) to (7). Robustness tests restrict the studied samples to forecasts close to the CFOs' median expected returns. The distance required in the different columns ranges from 1% to 10%.

### Table E.1. Learning Hypothesis: Limited to Expected Returns Near Consensus

This table shows estimates from ordinary least squares regressions of $\Delta \mathrm{CI}_{q-4\to q}$, the change in CFOs' confidence intervals from one forecast to the next (keeping four-quarter spacing). $\mathrm{I(Miss\ CI)}_{i,q-4}$ is a lagged indicator of whether the S&P 500 returns missed the previous CI. $\mathrm{CI}_{q-4}$ is the previous CI. Sample observations are limited to those in specified proximity to the median expected return across CFOs during a given survey date. Standard errors are double-clustered by the survey's calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | $\Delta \mathrm{CI}_{q-4\to q}$ (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\mathrm{I(Miss\ CI)}_{q-4}$ | 0.006 | −0.470 | −0.768 | −0.797 | 0.079 | −0.423 | −0.552 | −0.614 |
| | (0.771) | (0.529) | (0.548) | (0.494) | (0.820) | (0.571) | (0.657) | (0.589) |
| Survey date FE | No | No | No | No | No | No | No | No |
| Forecaster FE | No | No | No | No | No | No | No | No |
| Survey date FE × CI | Exact | Exact | Exact | Exact | 2pp | 2pp | 2pp | 2pp |
| abs(Exp ret − median) | 1pp | 3pp | 5pp | 10pp | 1pp | 3pp | 5pp | 10pp |
| Observations | 1,783 | 3,644 | 4,613 | 5,188 | 1,764 | 3,647 | 4,620 | 5,210 |
| Adjusted $R^2$ | 0.174 | 0.213 | 0.195 | 0.196 | 0.148 | 0.213 | 0.191 | 0.193 |

# Appendix F Miscalibration and Life Experiences

To further explore the birth year effect discussed in Section 4.1, we follow the methodology developed in Malmendier and Nagel (2011) to measure the relation between life experiences and miscalibration. We construct weighted average lifetime market return volatility for each CFO, using data from their birth year until the survey response date. We then regress adjusted CI width on this measure of experienced market volatility. As in the Malmendier and Nagel analysis, the weights used to construct the average are part of the estimation process, which makes the entire procedure non-linear. We similarly construct weighted average lifetime economic uncertainty using the historical Economic Policy Uncertainty Index (EPU) of Bloom et al. (2017) updated through 2024. The estimates of this regression are presented in Table F.1

As in Malmendier and Nagel (2011), we interpret the coefficients by measuring the change in CI width by moving from the $10^{\text{th}}$ to $90^{\text{th}}$ percentile of lifetime experience. For realized volatility, increasing lifetime experienced return volatility from the $10^{\text{th}}$ to $90^{\text{th}}$ percentile increases adjusted CI width by 3.0pp. Similarly, increasing lifetime experienced economic policy uncertainty from the $10^{\text{th}}$ to $90^{\text{th}}$ percentile increases adjusted CI width by 3.1pp. Overall, these results are qualitatively consistent with the life experience mechanism, but the estimates are statistically insignificant.

In both cases, the estimated weight parameters are positive and less than one, consistent with the life experience story of Malmendier and Nagel (2011) as being a driver of CFO miscalibration. The weighting parameters imply that the weighting function is concave decreasing in the number of years since the CFO was born. To illustrate this, Figure F.1 plot the weighting function for a CFO who is 50 years old with a weighting parameter of 0.473. This figure shows the weight assigned to the realized volatility for each year of the CFO's life, from the year they were born until they reach age 50. For the sake of comparison, we also plot the weighting function with a weighting parameter of 1.325, which is the estimate in Table III of Malmendier and Nagel (2011) for their specification that regresses stock market

**Table F.1. Miscalibration and Lifetime Experiences**

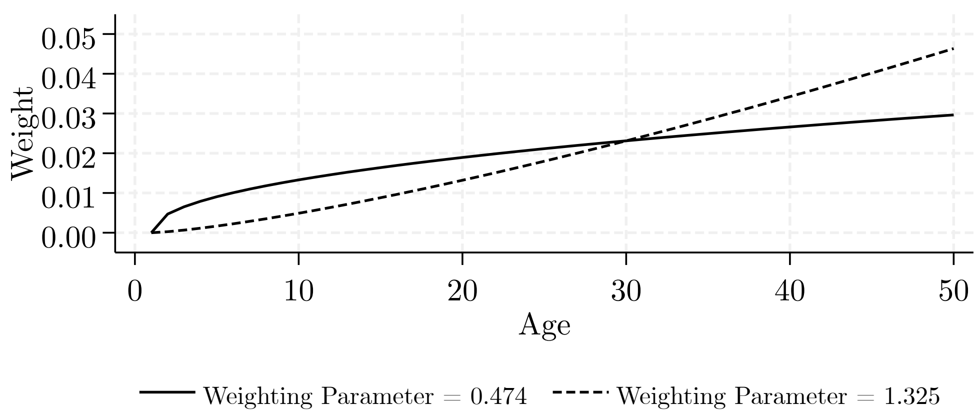
This table shows estimates of miscalibration on lifetime experiences of realized volatility and policy uncertainty as measured by the Economic Policy Uncertainty Index constructed by Bloom et al. (2017). In each column, the coefficient on experienced volatility or EPU and the weighting parameter are jointly estimated. Standard errors are double-clustered by the survey's calendar date and participant. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | CI (%; adjusted) | |
|---|---|---|
| | (1) | (2) |
| Experienced Volatility | 1.472 | |
| | (0.747) | |
| Experienced Economic Policy Uncertainty | | 0.126 |
| | | (0.056) |
| Weighting Parameter | 0.473 | 0.694 |
| Observations | 4,964 | 4,964 |
| Adjusted $R^2$ | 0.008 | 0.010 |

specification on lifetime experienced stock returns.

**Figure F.1. Weighting Function for Average of Life Experiences**

This figure presents the weight of each age of a CFO's life for two weighting parameters. The first, 0.474, is estimated in our analysis for life experienced realized volatility. For the sake of comparison, the second, 1.325, is estimated in Malmendier and Nagel (2011) in the specification that regresses stock market participation on lifetime experienced stock returns.




For both values of the weighting parameter, there is relatively more weight on more recent observations. Our estimates place relatively more weight on younger years, which implies an even stronger lifetime experience mechanism.