THE PERSISTENCE OF MISCALIBRATION

Michael Boutros
Itzhak Ben-David
John R. Graham
Campbell R. Harvey
John W. Payne

The Persistence of Miscalibration
Michael Boutros, Itzhak Ben-David, John R. Graham, Campbell R. Harvey, and John W. Payne
NBER Working Paper No. 28010
October 2020
JEL No. D03,D83,D84,E03,G30,G41

ABSTRACT

Using 14,800 forecasts of one-year S&P 500 returns made by Chief Financial Officers over a 12-year period, we track the individual executives who provide multiple forecasts to study how their beliefs evolve dynamically. While CFOs' return forecasts are systematically unbiased, their confidence intervals are far too narrow, implying significant miscalibration. We find that when return realizations fall outside of ex-ante confidence intervals, CFOs' subsequent confidence intervals widen considerably. These results are consistent with a model of Bayesian learning which suggests that the evolution of beliefs should be impacted by return realizations. However, the magnitude of the updating is dampened by the strong conviction in beliefs inherent in the initial miscalibration and, as a result, miscalibration persists.

Michael Boutros
Duke University
Department of Economics
Durham, NC 27708
michael.boutros@duke.edu

Itzhak Ben-David
The Ohio State University
Fisher College of Business
606A Fisher Hall
Columbus, OH 43210-1144
and NBER
ben-david.1@osu.edu

John R. Graham
Duke University
Fuqua School of Business
100 Fuqua Drive
Durham, NC 27708-0120
and NBER
john.graham@duke.edu

Campbell R. Harvey
Duke University
Fuqua School of Business
Durham, NC 27708-0120
and NBER
cam.harvey@duke.edu

John W. Payne
Duke University
Fuqua School of Business
Durham, NC 27708-0120
jpayne@duke.edu

1. Introduction

Miscalibration of beliefs, defined as a form of overconfidence in which confidence intervals over uncertain outcomes are too narrow, has been documented in numerous psychological and economic studies. Laboratory examples 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). Ben-David, Graham, and Harvey (2013) (hereafter BGH) document severe miscalibration, finding that only 36% of realized stock market returns fall within the 80% confidence intervals provided by top executives.[1] Miscalibration affects the real economy when overconfident agents act sub-optimally (Marray et al., 2020); for example, Barrero (2020) demonstrates that miscalibrated business managers can decrease their firms' values, leading to a sizable welfare loss.

There are many proposed explanations for this type of overconfidence, ranging from cognitive issues linked to anchoring, confirmation bias, and insufficient adjustment, to a general failure of intuitive thinking (see, for example, Kahneman (2011) and Moore et al. (2016)), and it is unlikely that there is a single cause to this effect. One of the potential remedies to miscalibration is feedback. However, most studies of miscalibration measure private beliefs at a single point in time, stopping short of considering whether miscalibrated agents re-evaluate their beliefs and make adjustments in light of new information. As such, these studies provide little insight into the persistence of miscalibration and whether it dissipates over time. In the presence of severe miscalibration, do individuals learn and improve their calibration over time? We seek to answer this question in this paper.

We use a unique dataset in which financial executives provide their beliefs about the variance of stock market returns in the form of confidence intervals. Many executives provide multiple predictions over time, i.e., after having observed the realization of their prior forecast. This setting allows us to measure how these executives update their beliefs. Our main finding is that when the realized return falls outside an executive's confidence interval, their subsequent confidence interval widens. While the widening of intervals is economically and statistically significant, the size of the widening is insufficient to obtain proper calibration, and, as a result, miscalibration persists.

Our analysis proceeds in three parts. First, we establish that CFOs' beliefs of future S&P 500 returns are informed by the historical distribution of returns, information which is freely available to all. CFOs pay attention to the benchmark S&P 500 return since their companies are routinely evaluated relative to the overall market and it is unlikely that they

---

[1]Miscalibration is also known as overprecision. More generally, these concepts fall under the umbrella of overconfidence, defined by Moore and Schatz (2017) as "excessive faith that [the agent] knows the truth."

can explain their firm's stock performance without reference to broad market movements. Our database consists of 14,800 predictions of the one-year return on the S&P 500 from 2000 to 2018 for more than 2,800 forecasters.[2] Remarkably, four CFOs have made over 40 forecasts. The survey respondents also provide 80% confidence intervals for their forecasts, which gauge CFOs' beliefs of expected return volatility. We have more than 4,000 pairs of observations for which we can observe the initial confidence interval, the return realization, and subsequent change in the confidence interval.

Our results show that while CFOs' forecasts of the one-year S&P 500 return are unbiased, they are extremely miscalibrated. We find that the average CFO forecast of the return is approximately 5.0% and the average realized annual return is also 5.0% over the same period, suggesting a lack of systematic bias. On the other hand, the standard deviation of realized annual returns is 17.1%, while the average of CFOs' beliefs of the standard deviation is only 5.5%, less than one-third of the historical experience. Equivalently, the distance between the 90th and 10th percentiles of the historical distribution is 43.8 percentage points (pp) wide, while the average 80% confidence interval in our sample is only 14.3 pp wide. This difference is the miscalibration documented in BGH.

In the second part, we use our unique dataset to assess how CFOs' beliefs evolve in response to observed realizations of stock market returns. Our baseline result is that CFOs who miss their confidence interval widen their subsequent interval by almost 15%, but since their initial intervals are only 12.3 pp wide, this amounts to updating by only 2.0 pp. Given that our data elicits separate estimates of the upper and lower bounds of the confidence interval, we also examine whether or not CFOs asymmetrically adjust their confidence intervals in response to missing the interval from above or below. We find that CFOs who miss the interval on the downside adjust the lower parts of their intervals by relatively more than the upper parts, and, similarly, CFOs that miss on the upside adjust the upper parts of their intervals by relatively more than the lower parts.

We exploit the long panel of repeat forecasters to measure the relation between learning and forecasting experience and find that CFOs learn less with each subsequent miss of the interval. The first time a CFO misses the interval, they widen by 3.0 pp, 1.5 as large as the baseline estimate for all misses. By the ninth time a CFO misses, however, we estimate no widening of the CI. We also find that learning decreases with initial miscalibration, which may be why miscalibration persists. We group CFOs into quartiles by initial calibration

---

[2]Our paper adds to the growing literature studying expectations held by firms and their executive officers. See, for example, Malmendier and Tate (2015), Bloom et al. (2017), Greenwood and Shleifer (2014), and Gennaioli, Ma, and Shleifer (2016).

measured over their first four forecasts and find that CFOs initially most miscalibrated widen their subsequent CIs by only 3.2% in response to missing, less than one quarter of the baseline estimate for all CFOs. On the other hand, CFOs initially least miscalibrated widen their subsequent CIs by 23.7% in response to missing, approximately 50% larger than the baseline estimate.

In the third part, we show that these results are broadly in line with an agent who uses Bayes rule to update their beliefs about the unknown variance of a stochastic return process.[3] Relative to the existing literature, our contribution is modeling the evolution of beliefs of the variance of the unknown process. The model has three key predictions that are borne out in the data.

First, if the observed return falls outside of a confidence interval constructed using the forecaster's prior belief of the variance, then the new posterior belief of the variance is larger, implying a wider subsequent confidence interval. Second, with each observed return and new posterior belief, the tightness or strength of the belief increases. As the forecaster observes new returns and forms new beliefs, they increasingly become more certain of their beliefs, and this additional experience leads to less updating in response to news. Third, the magnitude of updating crucially depends on not only the size of the miss, but also on the strength of the prior belief, which is tied to the degree of miscalibration. Forecasters who are initially more miscalibrated, or strongly hold their beliefs, update less in response to news.

Taken together, our findings characterize the evolution and persistence of CFOs' miscalibration: while confidence intervals widen with repeated misses, the magnitudes of the changes are small, decrease with experience, and decrease with initial miscalibration. CFOs learn in a manner consistent with Bayesian learning, but not by enough to attain proper calibration, and this is especially true for those initially most miscalibrated. As a result, miscalibration persists.

These results contribute to the literature seeking to understand over- and under-reaction of economic agents to news.[4] Most related to this paper is Bordalo et al. (2020), which studies revisions of stock return forecasts and finds that at shorter horizons, i.e. one-year-ahead, forecasters under-react to news (see Section III.B of the Internet Appendix). Our model and empirical findings suggest that this under-reaction may be because of how strongly CFOs

---

[3]A growing literature documents heterogeneity in the formation and evolution of beliefs and the impact of this heterogeneity on outcomes. See, for example, Kuchler and Zafar (2017), Fermand et al. (2019), Giglio et al. (2019), Meeuwis et al. (2018), Gervais and Odean (2001), Kuhnen (2014), and Martin and Papadimitriou (2020).

[4]See, for example, Shiller (1981), Abarbanell and Bernard (1992), Ali et al. (1992), Gennaioli and Shleifer (2010), and Bouchaud et al. (2019).

hold their initial beliefs.

Section 2 describes our data and documents that CFO beliefs are miscalibrated. Section 3 discusses CFOs' prior beliefs and their relation to the historical distribution of returns and, in Section 6, we develop our model of Bayesian learning. The main evidence of recalibration and learning is presented in Sections 4 and 5. Some concluding remarks are offered in the final section.

2. Data

We use a set of stock market predictions made by financial executives in the Duke-CFO survey. Each quarter, the survey is electronically delivered to senior financial executives and subscribers of *CFO* magazine. The survey contains a set of questions that appear in every survey and several topical questions which poll CFOs on important events related to current economic and geopolitical conditions. This dataset has been used in several prior academic studies. BGH study miscalibration patterns for CFOs. Greenwood and Shleifer (2014) analyze a long time series of investor expectations and their relation to expected returns in standard finance models. Gennaioli, Ma, and Shleifer (2016) look at a sample of firms and analyze the relation between expectations of growth and investment. In contrast to the learning mechanism developed in this paper, they argue that firms often use simple extrapolations in forming their next-period beliefs of growth.

The primary survey question in which we are interested asks:

> *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 __%.*

This question elicits both a point estimate of the mean as well as an 80% confidence interval for future realized returns.[5] These inputs can be used to calculate an imputed standard

---

[5]Bloom et al. (2017) use data from the Census Bureau's Management and Organizational Practices Survey on firms' reported subjective probability distributions of important future outcomes such as employment and input costs. They find that firms' subjective expectations are generally coherent probability distributions and similar to historical data, indicating that firms are able to generate accurate subjective distributions based on previous observations of data. Similarly, the evidence presented in the next section of CFOs' unbiased forecasts and relation to historical data indicate that they understand and coherently respond to the questions on our survey.

deviation for every forecast.[6] BGH study S&P 500 return predictions collected over 40 quarterly surveys between 2001Q2 and 2011Q2. We update this database to include an additional 24 surveys from 2011Q3 to 2017Q3. Our full sample has over 24,000 individual observations, almost 11,000 more than in BGH. In total, we have 14,800 responses for which we can identify the respondent and thus track the evolution of their predictions. Figure 1 illustrates that we are able to construct a large panel by tracking respondents over time. Almost 1,000 executives have responded to the survey exactly twice. Over 400 respondents have responded to the survey at least nine times, and there are almost two dozen CFOs who have responded more than 30 times.

Figure 1: Overview of Panel Length
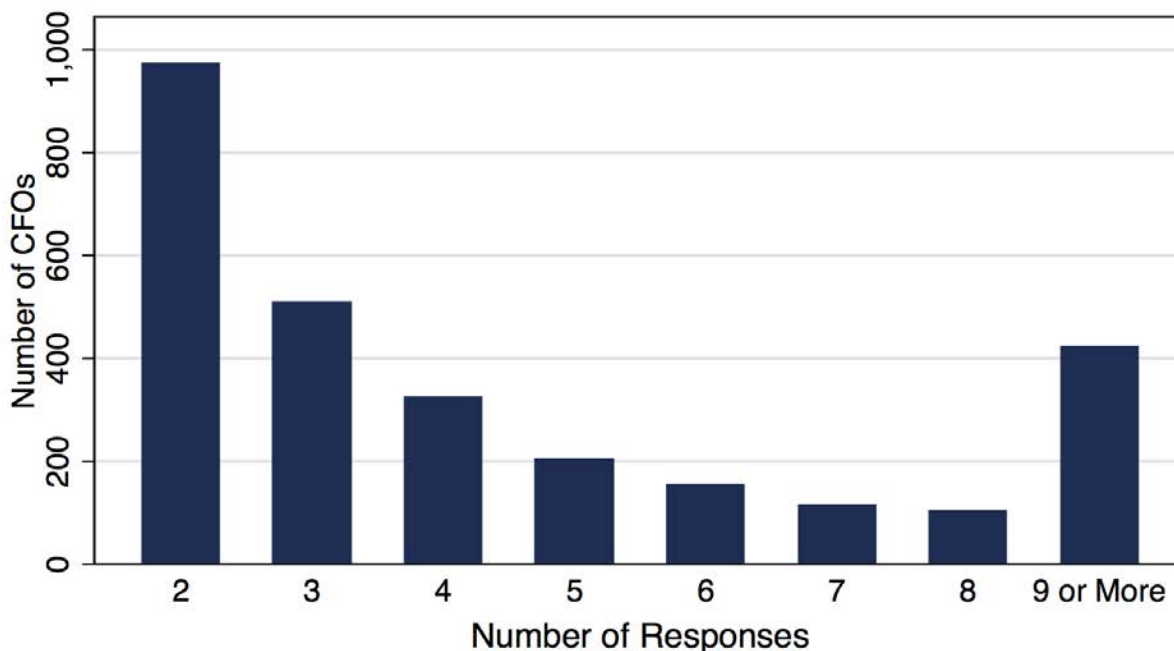

Figure shows the number of forecasts given by CFOs in our sample. For example, approximately 950 CFOs responded to the survey exactly twice.

---

[6]Keefer and Bodily (1983) show that the following method for calculating the imputed standard deviation of a continuous random variable, given the 10th and 90th percentiles, is preferred:

$$\sigma = \frac{P_{90} - P_{10}}{2.65}.$$

2.1. Managerial Miscalibration

In the BGH sample, CFOs hit their 80% confidence intervals only 36.8% percent of the time, providing striking evidence of miscalibration.[7] Our updated data provide an opportunity for an out-of-sample test. Interestingly, since 2011Q3, CFOs only hit their confidence intervals 24.0% of the time, suggesting they have become *more* miscalibrated on average. This difference of 12.8% is both economically and statistically significant.

Figure 2: Time Series of CFO Calibration
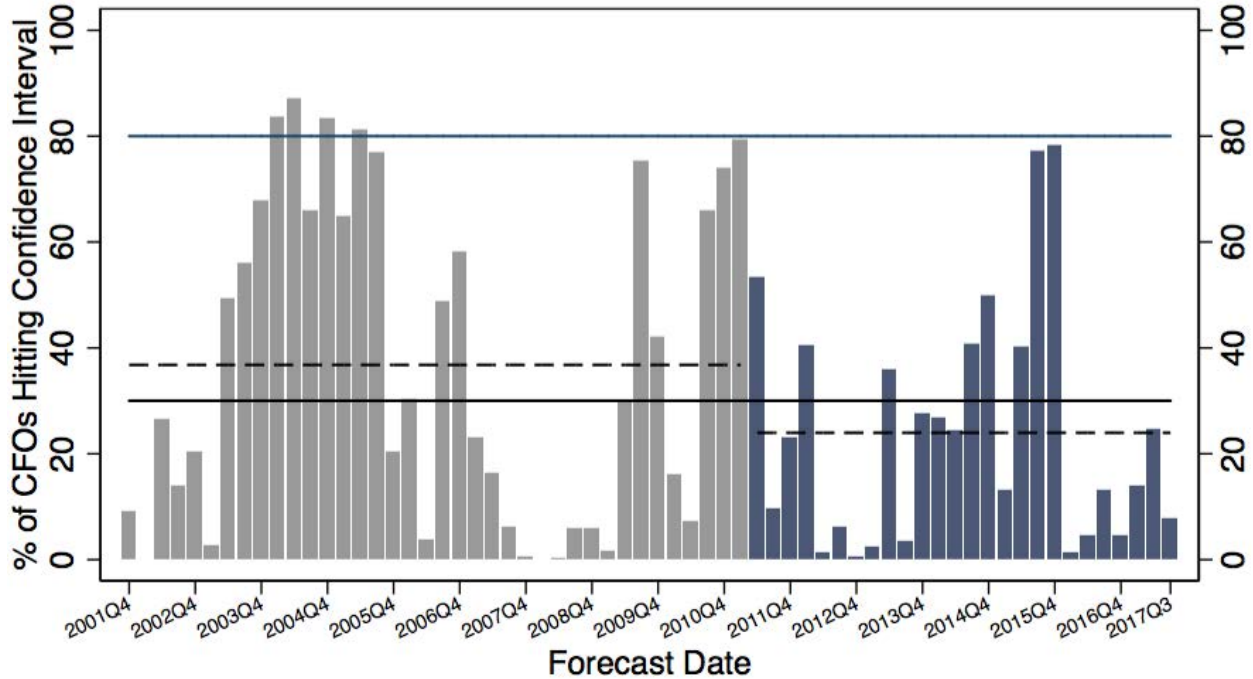
The percentage of survey responses with the ex-post return falling within the ex-ante 80% confidence interval (calculated using All Forecasts, see Table 1). The grey bars represent the sample covered by Ben-David, Graham, and Harvey (2013) and the blue bars are the new survey periods in our sample. The solid black line is the sample average calibration across all surveys and each dashed black line is the subsample average. The average number of respondents per survey is 355 in the earlier part of the sample and 395 in the later part of the sample.

Over the full sample, CFOs hit their 80% confidence intervals only 31.5% of the time. Figure 2 illustrates the percentage of responses in each survey that hit the forecast interval. We calculate annual S&P 500 returns in 12-month rolling windows from 1950 to 2018. The 10th and 90th percentiles are -12.5% and 28.1%, respectively, implying that an 80% confidence interval is 40.6 pp wide. Only 3.5% of responses have confidence intervals at least 40.6

---

[7]In BGH, CFOs hit their 80% confidence intervals over the first part of the sample, from 2001Q2 to 2011Q2, 36.3% of the time. This minor difference is due to a slightly modified algorithm for cleaning and merging data across surveys.

pp wide. These results using the full sample are consistent with those in BGH. Their results appear to be validated out-of-sample, and, if anything, CFOs appear more miscalibrated than in the original study. In our analysis, we take the initial miscalibration as given and study whether forecasters are able to attain proper calibration through learning. Although our data does not allow us to further investigate why CFOs are miscalibrated, there is a wide range of plausible explanations, from life experiences (Malmendier et al., 2011) to genetic heritability (Cesarini et al., 2009).

Table 1: Confidence Interval Width Summary Statistics

	All Forecasts			Single Forecasts			Repeat Forecasts		
	Hit	Miss	All	Hit	Miss	All	Hit	Miss	All
Mean	20.7	11.7	14.3	19.8	11.2	13.7	22.9	12.8	15.6
Median	18.0	9.0	10.0	15.0	8.0	10.0	20.0	10.0	12.0
25th Percentile	10.0	5.0	6.0	10.0	5.0	6.0	12.0	5.0	6.0
75th Percentile	30.0	15.0	20.0	25.0	15.0	20.0	30.0	17.0	20.0
Observations	4,308	10,492	14,800	3,015	7,142	10,157	1,293	3,350	4,643

Summary statistics of 80% confidence interval widths. Repeat Forecasts are those for which we have the respondent's initial forecast, the realized return, and the subsequent forecast. The reported values are for the subsequent confidence interval. Single Forecasts are those for which we do not observe the subsequent forecast. The forecast hit the confidence interval if the observed return falls within the confidence interval and missed otherwise. The 80% confidence interval for annual S&P500 returns in 12-month rolling windows from January 1950 to September 2018 is 40.6 percentage points wide.

2.2. Width of Confidence Intervals

In this and the next section, we present simple summary statistics of the data used in our primarily analysis. Table 1 presents statistics about the widths of confidence intervals in our sample. The first three columns represent the entire sample of forecasts. The mean width of CFO 10th to 90th percentile confidence intervals for market returns is 14.3%. As noted above, the distribution of annual S&P 500 returns implies an 80% confidence interval that is 40.6% wide. The 75th percentile of forecasters have a confidence interval (CI) width of 20.0%, which is still only half of the interval implied by the actual return data. CFOs' confidence intervals are far too tight around their point estimates, indicating miscalibration.

Conditioning on whether the realized return falls within the forecast interval ex-post, we find that, not surprisingly, those who were successful in hitting the interval had ex-ante wider intervals. The median forecaster who hits the interval has an interval twice as wide as the forecaster who misses, and this pattern is consistent at other percentiles of the distribution.

To study learning, we analyze pairs of "Repeat Forecasts" from the same respondent where we observe the initial confidence interval, the realized return, and the new confidence

interval exactly four quarters ahead. The final three columns of Table 1 summarize the new confidence interval for 4,643 Repeat Forecasts, with the middle three columns summarizing the remaining 10,157 "Single Forecasts." Overall, the width of the CIs for the Repeat Forecasts is only slightly larger than the Single Forecasts, suggesting two preliminary findings. First, if there is learning, it appears limited. Second, the repeat forecasters are still badly miscalibrated, with a mean confidence interval of 15.6%, much smaller than the 80% interval based on historical S&P 500 returns of 40.6%.

2.3. Symmetry of Confidence Intervals

Our unique dataset separately elicits the upper and lower bounds of CFOs' 80% confidence intervals and we find that most CFOs' intervals are somewhat asymmetric. We construct a simple measure of asymmetry as:

$$A = \frac{r_U + r_L}{2} - r_P, \tag{1}$$

where r_P is the point forecast, r_U is the upper bound, and r_L is the lower bound. When $A = 0$, the interval is symmetric. When $A > 0$, the interval is skewed to the right. When $A < 0$, the interval is skewed to the left.

Figure 3: Distribution of Asymmetry in Full Sample
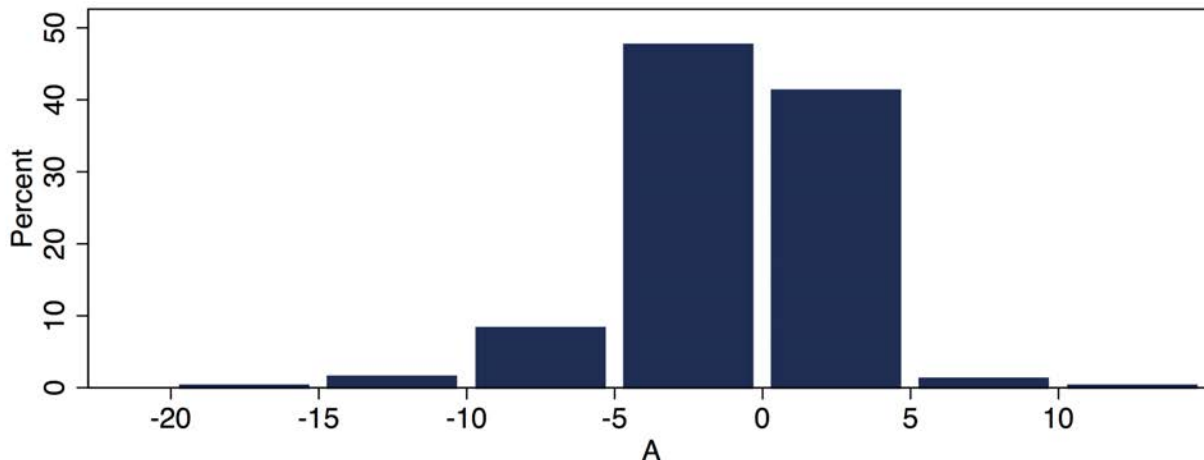
The distribution of the measure of asymmetry, $A = \frac{r_U + r_L}{2} - r_P$, where r_P is the point forecast, r_U is the upper bound, and r_L is the lower bound. When $A = 0$, the interval is symmetric, when $A > 0$, the interval is skewed to the right, and when $A < 0$, the interval is skewed to the left. We exclude the top and bottom percentile of observations.

Figure 3 presents the distribution of this measure of asymmetry. The mean asymmetry is approximately -1.5 pp and there are many more left-skewed confidence intervals than right-skewed. However, approximately 21% of the observations have zero asymmetry, and just

under 75% are symmetric to within two percentage points. Given this, it is not unreasonable to initially examine a model that imposes symmetry, though in our empirical analysis we exploit and analyze the asymmetries in confidence intervals.

2.4. Volatility

CFO confidence intervals are likely influenced by perceptions of market volatility. Given the well documented time-varying nature of volatility, if CFOs miss the interval because of an unpredictable change in volatility, characterizing the miss as miscalibration may be unfair. For example, in the 2008Q1 survey, none of the 280 CFOs interviewed hit their CI for the one-year-ahead S&P 500 return (-47.2%). Unexpectedly volatile markets in the wake of the financial crisis surprised CFOs to such a degree that both their point forecasts and confidence intervals were completely off the mark.

To understand how confidence intervals evolve, we study how forecasters' beliefs about volatility evolve through time. Distinguishing between expected and unexpected changes in volatility is important. Since CFOs' forecast intervals should incorporate their beliefs of expected changes in market volatility, unexpected volatility changes can render forecasted confidence intervals ex-post inaccurate, though ex-ante they might be well-formed given the information available when the forecast was made. Therefore, in studying the evidence surrounding learning and the evolution of volatility beliefs, we must be careful to control for both expected and unexpected changes.

As such, we construct a model of expected volatility by estimating a quarterly AR(1) process directly on the series of realized volatility (RV), measured using one year of daily S&P 500 returns. To match the horizon of our survey question, we generate a four-quarter-ahead forecast as our measure of expected volatility over the next year. In a given period, the expected change in volatility is the forecast for volatility minus realized volatility, both from the previous period. Unexpected volatility is the difference between current period realized volatility and the forecast of current volatility from the previous period. The two series can be summarized as:

$$\text{Unexp. Vol}_t = \text{Vol}_t - E_{t-1}\text{Vol}_t,$$
$$\text{Exp. } \Delta \text{ Vol}_t = E_{t-1}\text{Vol}_t - \text{Vol}_{t-1}.$$

The correlation between unexpected changes in volatility and forecasting accuracy, measured as the percentage of CFOs who hit the interval for each survey, is -0.29.[8] Unsurprisingly, this implies that when unexpected volatility is large, forecasting accuracy suffers.

---

[8]We also estimate a GARCH(1,1) model and find similar results.

3. Where Do Beliefs Come From?

Our ultimate goal is to study the evolution of beliefs and how CFOs update their expectations in response to observing new information. Thus far, we have documented severe miscalibration in CFO beliefs, updating the findings in BGH (2013). We now begin developing a nonparametric framework through which we study the dynamics of beliefs for this and the next two sections. In this section, we focus on how beliefs of the mean and variance are formed. We then explore how agents update their beliefs of the variance and study whether learning reduces miscalibration over time.

Later, in Section 6, we parameterize the model using standard distributional assumptions for the formation of beliefs and Bayes' rule for the updating of beliefs. Through the lens of the model, we bring together our various findings to present a collective stream of evidence that despite learning, miscalibration persists. To be clear, we do not argue nor formally test that CFOs literally use Bayes' rule to calculate their posterior belief distributions. Rather, the model provides a framework to guide our interpretation of the data.

### 3.1. How Do Forecasters Form Their Beliefs?

Figure 4 presents a process model of how each forecaster forms their belief of the return distribution. We assume that forecasters believe returns are normally distributed, and thus form beliefs over the two parameters that fully characterize the distribution: the mean, $\bar{r}$, and variance, σ_r^2. In the figure, the black captions describe the general process model and the blue captions detail a standard Bayesian parameterization of the process model. We return to the Bayesian model in Section 6 and focus in this section on the general model.

Our main survey question elicits a CFO's belief of 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.[9] 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.[10]

---

[9]We refer to this as Bayesian uncertainty and compare it to Knightian uncertainty in Appendix A.

[10]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 4: Process Model: Formation of Belief of Return Distribution
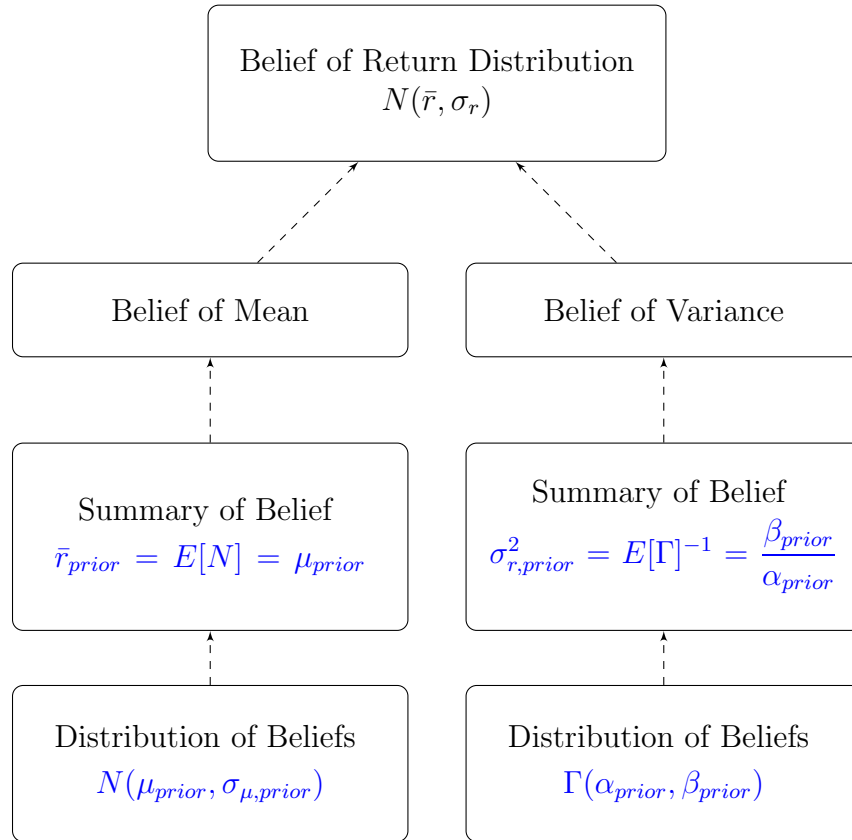
This figure illustrates the relation between the belief distributions of the mean and variance and the belief of the return distribution. Beginning from the lower left panel, the unknown belief of the mean is a normally distributed random variable. Moving up one panel, this distribution is summarized by its first moment, which forms the belief of the mean. In the lower right panel, the unknown belief of the variance is a gamma-distributed random variable, again summarized by its first moment. Together, these two beliefs form a belief of the return distribution, summarized in the topmost panel.

Focusing first on the left side of Figure 4, the survey question regarding the point forecast elicits the belief of the mean, $\bar{r}_{prior}$. Similarly, in the right side of the figure, the questions about the confidence interval elicit the belief of the standard deviation, $\sigma_{r,prior}$. These are the first moments of each belief distribution and are sufficient to generate forecasts of the future S&P 500 return distribution.

Where do these beliefs come from? CFOs likely use historical return distributions to generate the beliefs that they report in our survey. Suppose the CFO responds to the survey in 2018 and has access to stock returns dating back to 2000. Denote the series of N_r observations of historical returns by $\{r_i\}_{i=1}^{N_r}$. She wishes to use these data to generate beliefs of the mean and variance of the return distribution.

The natural candidate for the belief of the mean is the sample mean of the historical returns, and, similarly, the natural candidate for the belief of the variance is the sample variance. Table 2 compares the beliefs implied by CFOs responding to our survey to the historical distribution of realized returns. The first column calculates the average forecasted return and the average of the imputed standard deviation for each CFO forecast. The remaining columns calculate the mean and variance of historical return distributions over different samples. Specifically, we calculate the quarterly one-year-ahead S&P 500 return. Between 2000 and 2018, roughly corresponding to our survey sample period, the annualized mean of the historical return distribution was 5.0% and CFOs' belief of the mean of the return distribution was also 5.0%. This implies that overall, CFOs' beliefs of the mean of the return distribution are unbiased.[11]

The standard deviation of the historical return distribution is around 17% for all samples. This is three times larger than the average imputed CFO standard deviation of 5.5%. The third row of Table 2 calculates the ratio between CFOs' beliefs of the standard deviation and the historical standard deviation, where $\sigma_{CFO} = \lambda\sigma_{historical}$, where λ represents the ratio of perceived to realized volatility. Overall, we see that forecasters' beliefs of the standard deviation are approximately one third of the sample standard deviation.[12] Thus the belief of the standard deviation (and variance) is very poorly calibrated. Figure 5 plots two normal distributions parameterized by the means and standard deviations of CFO beliefs and real-

---

[11]We note that we are comparing the unconditional mean of the historical return distribution to CFOs' conditional belief of the mean of the return distribution. Assuming the return distribution is stationary, the similarity of the unconditional mean and the conditional one-year-ahead forecast implies that information at the forecast time is not useful for forecasting one-year-ahead returns.

[12]Barrero (2020) 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.

Table 2: Mean and Standard Deviation of CFO Beliefs and Historical Return Distributions

	CFO Beliefs of Return Distribution	Historical Returns (to 2018)				
		2000	1990	1980	1970	1960
Mean	5.0	5.0	8.9	10.1	8.5	8.4
Standard Deviation	5.5	17.1	16.5	16.8	17.1	16.4
Std. Dev. Ratio (λ)	–	0.322	0.333	0.327	0.322	0.335

Notes: CFOs beliefs of the return distribution are calculated using their forecasts of the S&P500 return. The mean is the average forecast of future S&P500 returns and the standard deviation is the average imputed belief of volatility (measured using the formula in footnote 6). The mean and standard deviation of the historical distribution of realized returns are calculated beginning from each listed year through the end of 2018. The standard deviation ratio, λ, is the ratio of the CFO belief of the standard deviation, 5.5%, and the historical standard deviation.

ized returns from 2000-2018. This figure illustrates the notion of miscalibration highlighted by BGH and others. Even if forecasters can correctly gauge the (unconditional) mean of the return distribution, their belief of the 80% confidence interval is far too narrow relative to the true distribution, and as a result, realized returns frequently fall outside this interval.

3.3. Conviction of Prior Beliefs

We now turn to the tightness or conviction of the beliefs. These are the second moments of the distributions of beliefs about the mean and variance of the return distribution. What are natural candidates for the tightness of the belief distributions? Recall that the beliefs of the mean and variance are derived from the sample mean and variance (of the historical return distribution). Since these sample moments are both random variables, their variances serve as natural candidates for the tightness. In particular, the belief of 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 σ_r^2 is the (unknown) population variance. Similarly, the belief of the variance is given by the sample variance, and thus the tightness is given by the variance of the sample

Figure 5: Distributions of CFO Forecasts and Historical Returns
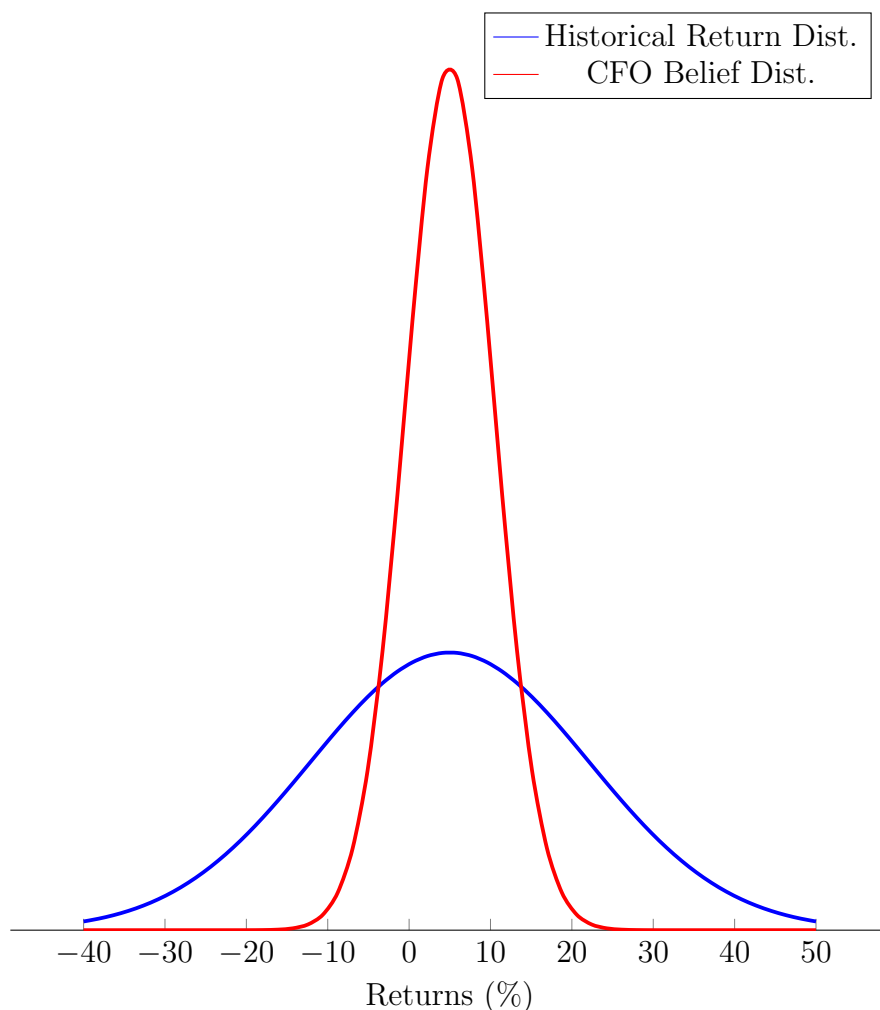
This figure presents the difference between the historical distribution of realized returns and CFOs' beliefs of the return distribution imputed from their forecasts. Though the data exhibit excess kurtosis and are not normally distributed, we illustrate these two distributions as normal distributions with mean and standard deviation calculated from the data (see Table 2). The blue line is parameterized by the mean and standard deviation of the historical return distribution between 2000 and 2018, which are, respectively, 5.0% and 17.1%. The red line represents CFOs' beliefs of the historical return distribution. The mean, 5.0%, is the average forecast of future S&P returns. The standard deviation, 5.5%, is the average of the imputed standard deviations, measured using individual CFOs' 80% confidence intervals and the formula in footnote 6.

Table 3: Summary of Belief Distributions and Sources of Beliefs

Unknown Parameter	Moment of Belief Distribution	Label	Plausible Source of Belief
Mean	mean	belief of mean	sample mean
	variance	conviction of belief	variance of sample mean
Variance	mean	belief of variance	sample variance
	variance	conviction of belief	variance of sample variance

variance,[13]

$$\frac{2\sigma_r^4}{N_r - 1}.$$

Since the population variance is unknown, it is estimated using the sample variance. Both of these variances are increasing in the sample variance. Altogether, then, the sample variance is used to inform the prior beliefs in three ways: the belief of the variance, and the tightness of beliefs about both the mean and the variance.

The analysis above highlights that CFOs' beliefs of the sample variance are much smaller than the historical data suggest they should be. As a result, since the variances of the belief distributions are simple increasing functions of the belief of the sample variance, it follows that CFOs' variances of the beliefs of the mean and variance are also much smaller than the data suggest they should be. We refer to this as CFOs having a strong conviction in their beliefs.[14] To be clear, miscalibration is when the mean of the belief of the variance is smaller (or larger) than what the realized data would suggest, while having a strong conviction is when the variances of the beliefs of the mean and variance are smaller than what the historical data suggest.

3.4. Summary

Table 3 summarizes the belief distributions and their sources. In our framework, CFOs' have prior beliefs over two distinct parameters: the mean of the return distribution and the

---

[13]See (Mood et al., 1974, p. 229) for a detailed derivation of the general case. For ease of exposition, presented here is the case where returns are normally distributed. 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 κ is defined as excess kurtosis.

[14]In the Bayesian literature, this is referred to as having strong, or tight, priors.

variance of the return distribution. In turn, a distribution of beliefs is formed for each parameter, and each of these distributions is characterized by a belief (i.e., mean) and tightness (i.e., variance).

The survey directly captures CFOs' beliefs of the mean and variance. We find that the belief of the mean is in line with the historical distribution of returns. However, consistent with the literature on miscalibration, we find that the belief of the variance is much smaller than what the historical data suggest. Although our data do not provide a direct measure of the tightness of CFOs' beliefs, as we showed, there exists a simple relationship between tightness of beliefs and the belief of the variance, and from this it follows that their beliefs are held with very strong conviction.

4. Learning and Recalibration

The framework developed in the previous section suggests that forecasters' beliefs are based on the historical distribution of returns. Given this, to study the dynamics of beliefs, we should explore how beliefs are updated in response to newly observed realized returns. Specifically, the empirical analysis in this section studies how forecasters update their beliefs of the variance in response to realized S&P500 returns. We estimate regressions of the form

$$\Delta \text{CI}_{it} = \alpha + \beta \cdot \text{Miss CI}_{it} + \gamma_i + \omega_t + u_{it}, \tag{2}$$

where ΔCI_{it} is the level change in CI width for forecaster i between time t and $t-1$, and Miss CI_{it} is an indicator that activates when the realized S&P 500 return at time t misses the CFO's interval produced at time $t-1$. The coefficient β measures the additional change in CI width for a forecaster who misses the interval relative to one who hits it. Forecaster and survey-quarter fixed effects are represented by γ_i and ω_t, respectively. Later, we also estimate specifications using changes in the upper and lower parts of the CI and using indicators that activate when the forecaster misses the interval from above or below.

In each set of regressions, we estimate different specifications of the model's control variables. We begin by estimating the model with no control variables, then add individual forecaster and fixed time effects. One important example of time fixed effects is unexpected return volatility, which we assume is the same for all forecasters at each survey date. Intuitively, higher unexpected volatility might increase beliefs of future volatility, which would widen a forecaster's confidence interval regardless of whether or not the previous forecast hit the interval. Therefore, in our third specification, we replace the general time fixed effects with the volatility controls described in Section 2.4.

Table 4: The Impact of Missing the Confidence Interval on Confidence Interval Widths

	Δ CI Width			
	(1)	(2)	(3)	(4)
Indicator: Miss CI	4.35***	5.21***	7.52***	4.76***
	(0.36)	(0.50)	(0.61)	(0.48)
Unexpected Vol.				0.15***
				(0.02)
Exp. Change in Vol.				0.16**
				(0.08)
Total Δ CI Width	1.61***	1.88***	1.99***	1.75***
	(0.16)	(0.16)	(0.25)	(0.16)
Total %Δ CI Width	13.0%	15.2%	16.2%	14.2%
Forecaster Fixed Effects	N	Y	Y	Y
Time Fixed Effects	N	N	Y	N
Volatility Controls	N	N	N	Y
Observations	4,643	4,643	4,643	4,643
R^2	0.035	0.041	0.126	0.068

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in CI width on indicator that activates when the realized return missed the confidence interval. Total Δ CI Width is the total change in CI width for a forecaster that misses the CI, calculated by summing the constant, fixed forecaster effects, fixed survey effects, and coefficient on the indicator. Total %Δ CI Width is the estimated total change in CI width divided into the average initial CI for all forecasters who missed the interval (12.3%). Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

4.1. Hitting or Missing the Confidence Intervals

Table 4 presents estimates of the baseline regression in (2), i.e., from regressing changes in CI width on an indicator that activates when forecasters miss their confidence intervals. In the first column, with no controls, the estimated effect of missing the interval is a widening of the subsequent CI by a significant 4.4 pp more than a forecaster who hits the interval. Adding forecaster fixed effects in the second column increases this estimate slightly to 5.2 pp.

Controlling for both forecaster and time fixed effects in the third column, we find that a forecaster who misses the CI widens their subsequent interval by 7.5 pp more than a forecaster who hits the CI. This is our baseline result and it is statistically significant at less than the 1% level. To calculate the total change in CI width for a forecaster who hits the interval, we sum the constant, average forecaster fixed effects, and average time fixed effects. To then calculate the total change for a forecaster who misses the interval, we add the coefficient on the indicator, β. Across all forecasters and surveys, the total change in CI width for the average forecaster who hits the CI is approximately -5.6 pp, i.e., a narrowing of the CI. The average forecaster who misses the CI instead widens their interval by approximately 2.0 pp. The average initial CI width among forecasters who missed is 12.3%, and thus the percentage change for forecasters who missed is approximately 16.2%. This estimate is both statistically and economically significant.

Replacing general time fixed effects with only volatility controls in the fourth column, the estimated difference between forecasters who miss relative to those who hit decreases to 4.8 pp. Interestingly, the estimate from controlling only for volatility is very similar to the estimate with no time fixed effects, but smaller in magnitude, implying that the time fixed effects are capturing more than just our estimated volatility effects. The fact that the common volatility controls produce half the explanatory power emphasizes the need for individual fixed effects to capture additional heterogeneity in beliefs about expected volatility.

4.2. Asymmetries: Missing High vs. Low and Upper vs. Lower CI

Table 5 presents estimates from regressing the total change in CI width on indicators that activate when the forecaster misses their interval high or low. A forecaster misses the interval high when the realized return is greater than their reported upper bound, and misses the interval low when the realized return is less than the forecaster's reported lower bound. Specifically, we estimate the regression

$$\Delta \text{CI}_{it} = \alpha + \beta_H \cdot \text{Miss CI High}_{it} + \beta_L \cdot \text{Miss CI Low}_{it} + \gamma_i + \omega_t + u_{it}, \qquad (3)$$

where, as in (2), $\Delta\mathrm{CI}_{it}$ is the level change in CI width for forecaster i between time t and $t-1$, forecaster fixed effects are γ_i, and time fixed effects are ω_t. The indicator Miss CI High$_{it}$ activates when the realized S&P 500 return at time t is larger then the upper bound of the CFO's interval produced at time $t-1$, and Miss CI Low$_{it}$ activates when the realized return is less than the lower bound of the CFO's interval. If the realized return falls within the interval, both indicators are equal to zero. The coefficients β_H and β_L measure the additional change in CI width for a forecaster who misses the interval high or low, respectively, relative to a forecaster who hits the interval. Again, measuring the total change in CI width for a forecaster requires summing the constant, forecaster fixed effects, time fixed effects, and the coefficient on the indicator. For brevity, Table 5 includes only three estimated specifications of the model: no fixed effects, individual forecaster and general time fixed effects, and individual forecaster and volatility fixed effects.

With no controls in the first column, forecasters who miss high widen by 3.2 pp more than those who hit the interval, while forecasters who miss low widen by 6.9 pp relative to those who hit. In total, forecasters who miss high widen by 0.4 pp, which is only 3.2% of the average initial CI width for forecasters who miss high (13.3%). On the other hand, forecasters who miss low widen by 4.1 pp in total, which is 39.8% of the average initial CI width for forecasters who miss low (10.4%). These estimates are all statistically significant.

The estimates with forecaster and time fixed effects tell a different story. Forecasters who miss high widen by 8.7 pp more than those who hit, and 3.0 pp in total, or almost 23% of the average initial width of forecasters who miss high. These estimates are significant at less than the 1% level. Forecasters who miss low widen by 5.5 pp more than those who hit, and this estimate is significant at the 1% level. However, the estimated total change in CI width for forecasters who miss low is not precisely estimated.

When general time fixed effects are replaced with volatility controls common to each forecaster, the estimates are similar to those in the first column with only forecaster fixed effects. Forecasters who miss high widen by 4.3 pp more than those who hit, and 1.1 pp in total, or approximately 8.4% of the average initial CI width of forecasters who miss high. Forecasters who miss low widen by 6.6 pp more than those who hit, and 3.5 pp in total, approximately one-third of the average initial CI width of forecasters who miss low.

Overall, the estimates in Table 5 make clear that forecasters who miss high widen their intervals differently than forecasters who miss low.[15] The regression with individual forecaster and general time fixed effects implies that forecasters who miss high widen more than

---

[15]In each regression, the coefficients on the miss high and miss low indicators are statistically different at less than the 1% level.

Table 5: The Impact of Missing the Interval High or Low on Confidence Interval Width

	Δ CI Width		
	(1)	(2)	(3)
Indicator: Miss CI High	3.17***	8.68***	4.25***
	(0.38)	(0.75)	(0.53)
Indicator: Miss CI Low	6.86***	5.46***	6.60***
	(0.47)	(0.81)	(0.66)
Unexpected Vol.			0.09***
			(0.03)
Exp. Change in Vol.			0.20**
			(0.08)
Miss CI High			
Total Δ CI Width	0.43*	3.01***	1.11***
	(0.20)	(0.43)	(0.26)
Total %Δ CI Width	3.2%	22.7%	8.4%
Miss CI Low			
Total Δ CI Width	4.12***	-0.21	3.45***
	(0.34)	(0.71)	(0.49)
Total %Δ CI Width	39.8%	-2.0%	33.3%
Forecaster Fixed Effects	N	Y	Y
Time Fixed Effects	N	Y	N
Volatility Controls	N	N	Y
Observations	4,643	4,643	4,643
R^2	0.053	0.129	0.071

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in CI width on indicators that activate when the forecast return misses the interval high or low. The forecaster missed the interval high if the realized return was higher than the upper bound of the confidence interval. If the realized return was lower than the lower bound of the interval, the forecaster missed the interval low. Total Δ CI Width is the total change in CI width for a forecaster that misses the CI, calculated by summing the constant, fixed forecaster effects, fixed survey effects, and coefficient on the indicator for missing high or low. Total %Δ CI Width is the estimated total change in CI width divided into the average initial CI for all forecasters who missed the interval high (13.3%) or low (10.4%). Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

those who miss low, while the other regressions imply the opposite.

To further explore this ambiguity, we analyze how the upper and lower portions of the interval change when forecasters hit or miss their CIs high or low. As detailed in Section 2.2, CFOs in our sample provide separate lower and upper bounds of their confidence intervals, allowing us to construct the upper portion of the CI as the distance between the upper bound of the CI and the point forecast of the return. Similarly, the lower portion of the interval is the distance between the point forecast and the lower bound of the CI. Since we are interested in analyzing the shape of the CI and whether it changes, we focus on changes in each part of the CI to ignore effects of the CI changing because of changes in the point forecast.[16]

The resulting estimates show that CFOs whose forecasts miss high widen the upper portion of the CI more than CFOs who miss low, and CFOs who miss low widen the lower portion of the CI more than CFOs who miss high. Table 6 presents estimates from the regressions

$$\Delta \text{ Upper CI}_{it} = \alpha + \beta_H \cdot \text{Miss CI High}_{it} + \beta_L \cdot \text{Miss CI Low}_{it} + \gamma_i + \omega_t + u_{it}, \qquad (4)$$

$$\Delta \text{ Lower CI}_{it} = \alpha + \beta_H \cdot \text{Miss CI High}_{it} + \beta_L \cdot \text{Miss CI Low}_{it} + \gamma_i + \omega_t + u_{it}, \qquad (5)$$

where Δ Upper CI$_{it}$ is the change in the upper portion of the CI (UCI) and Δ Lower CI$_{it}$ is the change in the lower portion of the CI (LCI). The coefficients β_H and β_L measure the additional change in UCI or LCI width for a forecaster who misses the interval high or low, respectively, relative to a forecaster who hits the interval. On average, the UCI was 5.0% wide for forecasters who missed high and 4.4% wide for forecasters who missed low, and the LCI was 8.3% wide for forecasters who missed high and 5.9% for forecasters who missed low.

Each pair of columns in Table 6 presents estimates of equations (4) and (5). In the first two columns, we do not control for fixed effects. In total, forecasters who miss high widen the UCI by approximately 4.8%, while the estimated total change in the LCI for forecasters who miss high is not statistically different from zero. Forecasters who miss low widen the UCI by 31.4% and widen the LCI by 46.0%.

In the third and fourth columns, we control for forecaster fixed effects and general time fixed effects. In total, forecasters who miss high widen the UCI by 38.2% and the LCI by 13.4%, while forecasters who miss low narrow the UCI by 30.2%. The estimated total change in the LCI for forecasters who miss low is not statistically different from zero. In the fifth

---

[16]Further results are shown in the appendix. Appendix C.1 presents results on changes in the upper and lower bounds of the CI when forecasters miss the interval, and Appendix C.2 presents results on changes in the upper and lower parts of the CI when forecasters miss the interval.

Table 6: The Impact of Missing the Interval High or Low on Upper and Lower Confidence Intervals

	Δ UCI (1)	Δ LCI (2)	Δ UCI (3)	Δ LCI (4)	Δ UCI (5)	Δ LCI (6)
Indicator: Miss CI High	1.42***	1.75***	4.01***	4.67***	1.96***	2.29***
	(0.19)	(0.30)	(0.36)	(0.60)	(0.26)	(0.42)
Indicator: Miss CI Low	2.57***	4.29***	0.78*	4.68***	2.02***	4.58***
	(0.24)	(0.33)	(0.41)	(0.65)	(0.31)	(0.51)
Unexpected Vol.					0.07***	0.02
					(0.02)	(0.02)
Exp. Change in Vol.					0.06*	0.13**
					(0.04)	(0.06)
Miss CI High						
Total Δ Width	0.24**	0.19	1.90***	1.11***	0.64***	0.46**
	(0.10)	(0.15)	(0.21)	(0.34)	(0.13)	(0.20)
Total %Δ Width	4.8%	2.3%	38.2%	13.4%	13.0%	5.6%
Miss CI Low						
Total Δ Width	1.39***	2.74***	-1.33***	1.13*	0.71***	2.75***
	(0.17)	(0.24)	(0.37)	(0.56)	(0.23)	(0.38)
Total %Δ Width	31.4%	46.0%	-30.2%	18.9%	16.0%	46.2%
Forecaster FE	N	N	Y	Y	Y	Y
Time FE	N	N	Y	Y	N	N
Vol. Controls	N	N	N	N	Y	Y
Observations	4,643	4,643	4,643	4,643	4,643	4,643
R^2	0.033	0.037	0.109	0.085	0.054	0.046

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in sizes of the upper and lower portions of the confidence interval on indicators that activate when the forecaster missed the interval high or low. The upper portion of the confidence interval (UCI) is the distance between the upper bound and the forecast. The lower portion of the confidence interval (LCI) is the distance between the forecast and the lower bound. Total Δ Width is the total change in UCI or LCI width for a forecaster that misses the CI, calculated by summing the constant, fixed forecaster effects, fixed survey effects, and coefficient on the indicator for missing high or low. Total %Δ Width is the estimated total change in UCI or LCI width divided into the average initial UCI or LCI for all forecasters who missed the interval high or low. On average, the UCI was 5.0% wide for forecasters who missed high and 4.4% wide for forecasters who missed low, and the LCI was 8.3% wide for forecasters who missed high and 5.9% for forecasters who missed low. Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

and sixth columns, we replace general time fixed effects with volatility controls. In this specification, forecasters who miss high widen the UCI by 13.0% and the LCI by 5.6%, while those who miss low widen the UCI by 16.0% and the LCI by 46.2%.

Overall, our results indicate that forecasters who miss high widen the upper interval more than the lower interval, while forecasters who miss low widen the lower interval more than the upper interval. This implies that upon missing high, and controlling for the possibility that the entire interval shifts upward, forecasters change the shape of their interval, increasing the rightward-skew of their interval and placing a larger probability on a larger return realization. Similarly, upon missing low, forecasters increase the leftward-skew of their interval, placing a larger probability on a smaller return realization.

This evidence suggests that forecasters are aware of and respond not only to whether they missed their intervals, but also to whether the observed return was above or below their interval. Forecasters shift more mass to the upper portion of the CI when missing high, and shift more mass to the lower portion of the CI when missing low. We interpret this as further evidence that is consistent with learning.

4.3. Path Dependence

In this section, we study whether forecasters display any path dependence in their updating. For example, if at time t the forecaster misses the interval and at time $t+1$ the forecaster hits the interval, path-independent learning would imply that their new confidence interval is the same as if they had hit the interval at time t and missed at time $t+1$. We focus on a subset of the main sample for which we observe two consecutive four-quarter-apart forecasts and return realizations. The change in CI width is measured using the second forecast and the initial CI width. With two consecutive forecasts, there are four possible paths: {hit, hit}, {hit, miss}, {miss, miss}, and {miss, hit}. To test for path dependence, we examine a subsample consisting only of forecasters who followed two paths, {hit, miss} and {miss, hit}, referring to each group as hit-miss and miss-hit forecasters, respectively.

We estimate the regression equation

$$\Delta^2 \text{CI}_{it} = \alpha + \beta \cdot \text{HM}_{it} + \gamma_i + \omega_t + u_{it}, \tag{6}$$

where $\Delta^2 \text{CI}_{it} = \text{CI}_{it} - \text{CI}_{i,t-2}$ is the level change in CI width for forecaster i between time t and $t-2$, forecaster fixed effects are γ_i, and time fixed effects are ω_t. The indicator HM_{it} activates for hit-miss forecasters, i.e., when the forecaster hit the interval with their forecast at time $t-1$ and missed the interval with their time t forecast. The coefficient β measures the additional change in CI width for a hit-miss forecaster relative to a miss-hit forecaster. If there is no path dependence, this coefficient is zero. Table 7 includes estimates of the model

Table 7: Testing for Path Dependence for {Hit,Miss} and {Miss,Hit} Sequences

	(1)	(2)	(3)
Difference	-2.12**	-5.18**	-1.35
	(1.05)	(2.32)	(1.28)
Hit, Miss	-0.95	-2.80**	-0.60
	(0.75)	(1.24)	(0.58)
Miss, Hit	1.17*	2.38*	0.75
	(0.73)	(1.33)	(0.74)
Forecaster FE	N	Y	Y
Time FE	N	Y	N
Vol. Controls	N	N	Y
Observations	708	708	708
R^2	0.007	0.258	0.039

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in CI width between two forecasts (i.e., t and $t-2$) on an indicator that activates for hit-miss forecasters (i.e., when the forecaster hit the interval with their forecast at time $t-1$ and missed the interval with their time t forecast). Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

with no fixed effects, individual forecaster and general time fixed effects, and individual forecaster and volatility fixed effects.

In all three specifications, there is an economically large difference between the miss-hit and hit-miss sequences. With forecaster and general time fixed effects, hit-miss forecasters have an interval that is approximately 5.2 pp narrower than miss-hit forecasters. Hit-miss forecasters narrow their confidence intervals over the two responses by 2.8 pp, while miss-hit forecasters widen their intervals by 2.4 pp. In Section 6.3.2, we explain how path dependence may occur in an environment with Bayesian learning due to relative conviction of belief that can vary in the ordering of events.

5. The Persistence of Miscalibration

Having established that CFOs appear to learn by widening their confidence intervals in response to newly observed returns, in this section we study whether the learning is sufficient to obtain proper calibration. Of the 4,643 forecasts matched with return realizations and new forecasts in our sample, approximately 70% miss the confidence interval. Of these, half of CFOs appear to learn and widen their subsequent confidence intervals. Since these CFOs forecast wider confidence intervals, they should be better calibrated. Indeed, we find that the group of CFOs that misses and learns hits the confidence interval just over 50% more often than the group of CFOs that misses and does not learn. However, both groups are still poorly calibrated: those who do not learn hit the confidence interval 19% of the time, while those who learn hit the confidence interval 29% of the time. We find that although CFOs learn, their confidence intervals are not widened by nearly enough to attain proper calibration, and thus miscalibration persists.

5.1. Recalibration and Initial Miscalibration

In this section, we test whether initial miscalibration predicts persistent miscalibration, even with learning. We divide forecasters into groups based on initial calibration and study how each group responds to missing the CI. For each CFO, we construct average CI width using their first four responses, and call this their initial CI width or initial calibration. We sort CFOs into three groups based on their initial calibration: $0-10$ pp, $10-20$ pp, and more than 20 pp. These groups roughly correspond to the lower quartile, interquartile, and upper quartile of CI widths (see Table 1).

Similar to the specification in (2), we estimate the regression equation

$$\Delta\text{CI}_{it} = \alpha + \sum_{j=1}^{3} \beta_j \cdot \text{Miss CI}_{it} \cdot \mathbb{1}\{\text{Initial Calibration Group} = j\} + \gamma_i + \omega_t + u_{it}, \qquad (7)$$

where the coefficient β_j measures the how much a forecaster in initial calibration group j widens their interval in response to missing relative to a forecaster that hits the interval. To estimate this regression, we exclude any of the forecasts used to construct the initial CFO calibration, reducing the sample size from 4,643 to 2,343 observations.[17] Table 8 presents the estimates from the regression with forecaster and time fixed effects. The first column re-estimates the baseline specification from Table 4 on the smaller subsample. In line with the earlier estimation, a forecaster who misses the interval widens their subsequent CI by approximately 2.0 pp or 11.7% of their initial CI width.

Forecasters initially most miscalibrated widen their CIs by 0.3 pp or 3.2% in response to missing the interval, and this estimate is not statistically different than zero. Forecasters in the middle group widen their CIs by approximately 1.5 pp or 8.4% and forecasters in the upper quartile of initial calibration widen their CIs by 7.4 pp or 23.7%. Overall, CFOs with the largest initial calibrations also respond the most to new information, widening their CIs by a statistically and economically large amount. CFOs who are initially the most miscalibrated not only begin furthest away from the correct calibration, but also update the least in response to missing. As a result, miscalibration persists.

5.2. Recalibration and Experience

We turn to analyzing learning over time as CFOs become more experienced with forecasting.[18] Similar to the specification in (2), we present estimations from the regression

$$\Delta\text{CI}_{it} = \alpha + \sum_{j=1}^{N} \beta_j \cdot \text{Miss CI}_{it} \cdot \mathbb{1}\{\text{Response \#} = j\} + \gamma_i + \omega_t + u_{it}, \qquad (8)$$
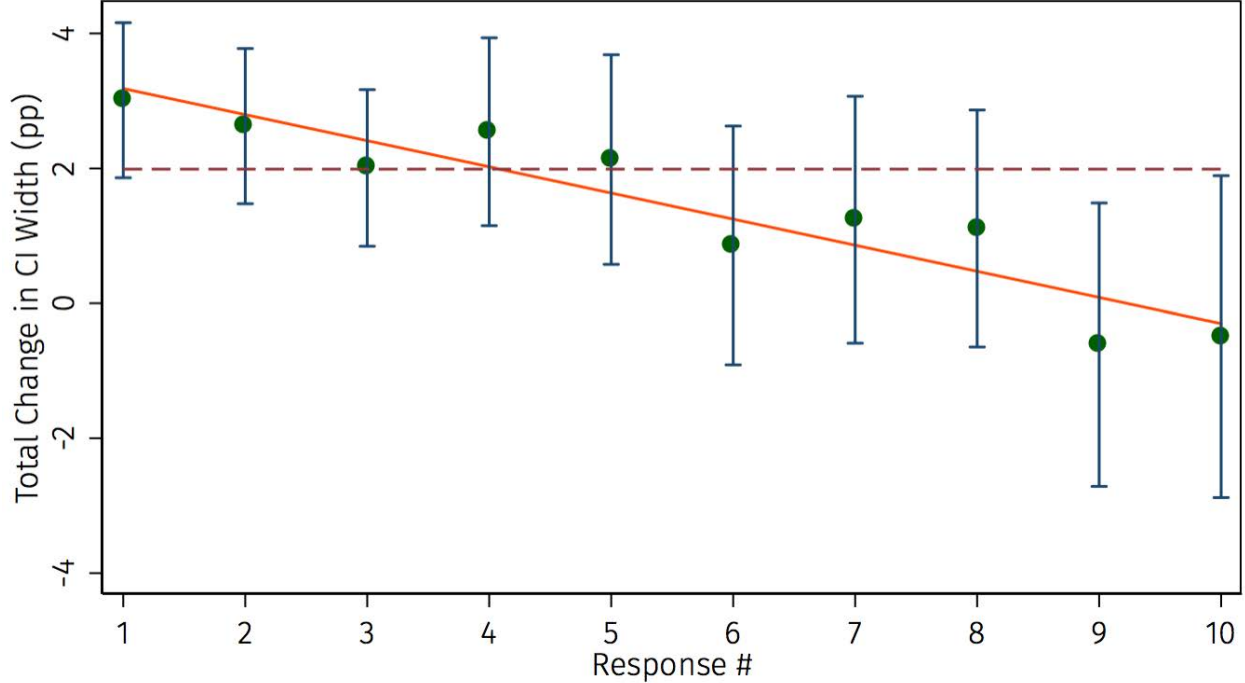
where the coefficient β_j measures the additional widening of the jth predicted CI for a forecaster who misses the interval.[19] Controlling for forecaster and general time fixed effects, Figure 6 plots the total change in the jth CI for a forecaster that misses the interval for responses one through ten. The dashed line is the baseline result in the second column of Table 4 that, on average, forecasters who miss the interval widen their subsequent CI by 1.99 pp in total.

---

[17]Recall that each observation consists of an initial forecast, the realized return, and the subsequent forecast. The subsample used in this analysis excludes any initial or subsequent forecast used to generate the initial CI width, and includes only forecasts by CFOs who have made at least four forecasts (which are used to calculate initial calibration).

[18]Murphy and Winkler (1977) find that weather forecasters have well-calibrated beliefs and suggest this is due to the frequent feedback on the accuracy of their predictions.

[19]Recall from Figure 1 that CFOs often respond to the survey multiple times, and that over 400 respondents have responded to the survey at least nine times.

Figure 6: Recalibration and Experience

This figure shows the total change in the jth predicted CI for a forecaster that misses the interval. Specifically, from (8), each point is the sum of the constant, average forecaster and time fixed effects, and the coefficient β_j, which represents the additional widening of the jth predicted CI. The dashed reference line is the baseline estimate for total change in CI width for all forecasts, 1.99 pp, corresponding to the second specification in Table 4.

As forecasters respond more often to the survey, they appear to learn less from each subsequent miss of their predicted intervals. Forecasters who miss the interval on their first response widen their intervals by approximately 3.0 pp, 50% more than our baseline estimation for all responses, and this estimate is statistically different than zero. By the ninth response, however, the learning has effectively vanished. These findings are consistent with the notion of CFOs learning from their mistakes but with diminishing returns to learning.

6. A Model of Bayesian Learning

To aid in cohesively interpreting the evidence we have presented thus far, we parameterize the process model in Figure 4 and, obtaining closed-form solutions, study the model's implications. To update their beliefs, forecasters use the Bayes' rule to generate a posterior distribution by combining their prior distribution with the newly observed return. As is standard in the Bayesian literature, we assume that the belief of the mean is normally distributed and the belief of the variance is distributed according to a gamma distribution. While the first moment of the belief distribution is sufficient to produce forecasts of future

Table 8: The Impact of Learning Based on Initial Calibration

| | (1) Baseline | (2) Interactions | | |
| | | Most Miscal. | | Least Miscal. |
	Miss CI	Miss CI × G1	Miss CI × G2	Miss CI × G3
Total Δ CI Width	1.97***	0.29	1.50*	7.35***
	(0.31)	(0.57)	(0.89)	(1.50)
Total %Δ CI Width	11.7%	3.2%	8.4%	23.7%
Observations	2,343		2,343	
R²	0.108		0.116	

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in CI width on an indicator that activates when the realized return misses the confidence interval. In (2), the indicator is interacted with a dummy variable for each of the initial calibration groups: $0 - 10$ pp, $10 - 20$ pp, and 20 pp or more. Total Δ CI Width is the total change in CI width for a forecaster that misses the CI, calculated by summing the constant, fixed forecaster effects, fixed survey effects, and coefficient on the indicator. Total %Δ CI Width is the estimated total change in CI width divided into the average initial CI for all forecasters in the group. Regressions include forecaster and time fixed effects.

returns, characterizing the second moment of the belief distribution, which describes the strength with which the belief is held, is important for understanding how beliefs evolve in response to newly observed information.

6.1. A Simple Model of Returns

We assume that forecasters believe returns are the sum of a constant mean, $\bar{r}$, and a serially uncorrelated white noise shock, $\epsilon_t \sim N(0, \sigma_r)$, such that

$$r_t = \bar{r} + \epsilon_t \sim N(\bar{r}, \sigma_r).$$

The forecaster's task is to generate a forecast of r_{t+1} using all information available up to and including r_t. This forecast will correspond with $\bar{r}$. Additionally, the agent must generate a confidence interval around her forecast, which will be a function of σ_r. The forecaster is Bayesian-uncertain about the values parameterizing the distribution of returns: $\bar{r}$ and σ_r. Bayesian updating of both unknown parameters of a Gaussian distribution is a well-known problem in Bayesian statistics. In the next two subsections, we analyze uncertainty about each parameter individually to isolate the primary channel through which the newly observed return impacts beliefs of the unknown mean and variance. In our empirical analysis of the belief of the variance, we use insights from the full model and allow the belief of the mean to change over time.

6.1.1. Unknown Mean, Known Variance

We first assume that the forecaster knows the variance of the return distribution, σ_r, but is uncertain about the mean, $\bar{r}$. This subsection focuses on the belief of the mean in the left side of Figure 4.[20] Starting from the bottom panel on the left side, the forecaster's initial belief of the unknown mean is a random variable distributed according to $N(\mu_{prior}, \sigma_{\mu,prior})$.[21] The distribution of beliefs about the unknown mean is summarized by its mean, so that

$$\bar{r}_{prior} = E[N(\mu_{prior}, \sigma_{\mu,prior})] = \mu_{prior}.$$

In the terminology developed in Section 3, μ_{prior} is the belief of the mean and $\sigma_{\mu,prior}$ is the strength of the belief of the mean. This variance is distinct from the belief of the variance of the return distribution, σ_r, though we argued earlier that the two are linked.

We choose a Gaussian prior distribution for three reasons. First, it is a commonly used and well-known distribution, and therefore suitable as a starting point. Second, the distribution is physically plausible since the unknown parameter can take on any negative or positive value. Third, when Bayes' rule is used to combine this prior with newly observed data, the resulting posterior distribution is also Gaussian. This property makes the Gaussian distribution a conjugate prior distribution. Conjugate priors are widely used in the Bayesian literature because they provide analytic tractability, allowing us to derive our main results in closed form.

When tasked with forming a belief for r_{t+1}, but before observing the current return, r_t, the forecaster's initial beliefs are given by $\bar{r}_{prior} = \mu_{prior}$. Figure 7 illustrates three examples of belief distributions with the same mean, μ_{prior}, but different variances. A forecaster with any of these distributions will have the same belief but different tightness or conviction in their belief of the mean. Thus any of these beliefs will generate the same point forecast; the differences in conviction of beliefs only become relevant when we ask how the CFO updates her beliefs upon observing new returns.

Suppose that now the CFO observes the return r_t and wants to form a new belief of the unknown mean, $\bar{r}_{posterior}$, given the initial belief, $\bar{r}_{prior}$. Combining the newly observed return with the prior belief of the mean yields a newly distributed belief $N(\mu_{posterior}, \sigma_{posterior})$, again summarized by $\bar{r}_{posterior} = \mu_{posterior}$. The well-known Bayesian updating rule yields that the

---

[20]Since the variance is known, there is no need to discuss the prior belief of the variance, i.e., the right side of Figure 4.

[21]In Figure 4, the black captions describe the general nonparametric process model, and the blue captions specify the parameterization developed in this section.

Figure 7: Examples of Belief Distributions of Unknown Mean
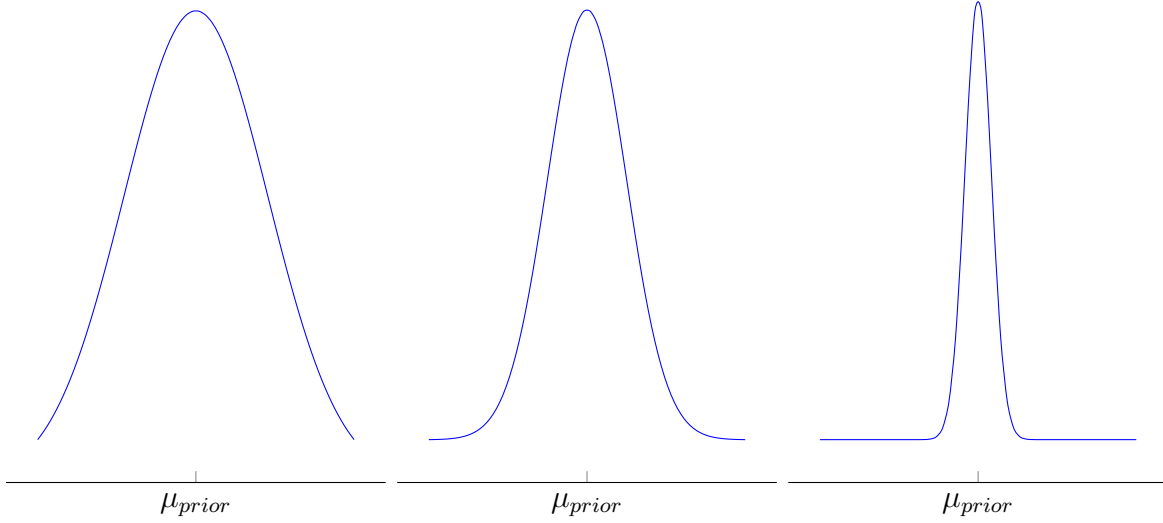
This figure illustrates three examples of normal distributions that represent beliefs of the unknown parameter, μ, the mean of the return distribution. All three have the same mean and represent the same belief: μ_{prior}. From left to right, the tightness or conviction in beliefs in the mean is increasing (i.e., variance is decreasing). Note that the tightness of the prior for the mean belief should not be confused with the narrow confidence intervals which are linked to the belief of the variance. In this example, we assume the variance is known.

new belief is directly related to the prior belief's mean and standard deviation:

$$\mu_{posterior} = \mu_{prior} + \frac{\sigma^2_{\mu,prior}}{\sigma^2_{\mu,prior} + \sigma^2_r}(r_t - \mu_{prior}).$$

The newly observed return enters in the second term of this expression as the difference between the observed return and the prior belief of the mean. This difference is multiplied by the ratio of the tightness of the belief about the mean and the (known) variance of the return distribution. As the CFO's conviction in her belief increases, this ratio decreases. Thus the more strongly held is the prior belief, the less will the CFO incorporate information from the newly observed return into her updated beliefs of the mean.

Consider two CFOs with the same initial beliefs, $\bar{r}_{prior}$, but different convictions in their beliefs. Both CFOs generate the same forecast for the S&P 500 return. However, upon observing the same realized return, r_t, their updated beliefs of the mean will be different because of the different convictions in their beliefs. Thus in a world with Bayesian learning, it is plausible to observe forecasters with similar initial beliefs diverging after observing the same public signal.

6.1.2. Known Mean, Unknown Variance

Next, we instead assume that the forecaster knows the mean of the return distribution, $\bar{r}$, but is uncertain about the variance, σ_r. Beginning in the lower right panel of Figure 4, we assume that the belief of the unknown variance is distributed according to $\sigma_{r,prior}^{-2} \sim \Gamma(\alpha_{prior}, \beta_{prior})$, and is again summarized by the mean of this distribution, so that

$$\sigma_{r,prior}^2 = E[\Gamma(\alpha_{prior}, \beta_{prior})]^{-1} = \frac{\beta_{prior}}{\alpha_{prior}}.$$

In the terminology developed in Section 3, $\sigma_{r,prior}^2$ is the belief of the variance, which is directly linked to the confidence intervals that we study.[22] We choose a gamma distribution since the variance must be strictly positive and because it is a conjugate prior distribution. The gamma distribution is parameterized by α, governing the shape, and β, which is called the inverse scale or rate parameter.[23] The shape parameter essentially determines the mass surrounding the peak of the probability density while the scale parameter governs thickness of the tails. As the value of the shape parameter increases, the peak of the pdf increases and more mass surrounds out. As the inverse scale decreases, the tails widen, drawing mass from the peak. The prior value for α can be interpreted as the strength of the prior; the larger is this parameter, the larger is the mass surrounding the peak.

The forecaster uses her belief of the variance of the return, $\sigma_{r,prior}^2$, to generate a confidence interval around her point forecast of the return. Figure 8 illustrates three examples of belief distributions that generate the same beliefs of the variance with different tightness or convictions. As above, each of these beliefs generates the same confidence interval, but different posterior beliefs in response to observing the same realized return.

Upon observing a new return, r_t, the forecaster's new belief of the variance is given by:

$$\sigma_{r,posterior}^2 = \frac{\beta_{posterior}}{\alpha_{posterior}} = \frac{\beta_{prior} + \frac{1}{2}(r_t - \bar{r})^2}{\alpha_{prior} + \frac{1}{2}},$$

where μ is the known mean. This updating expression is helpful for interpreting the hyperparameters of the gamma distribution. The first, α_{prior}, reflects the number of observations, while the second, β_{prior}, tracks the sum of squared differences between the mean of the return distribution and each observed return. The new value for α reflects the additional scaling required from using one more observation to inform the belief of the variance, while the

---

[22]As noted earlier, the tightness of the belief is given by the variance of the gamma distribution, $\frac{\beta_{prior}^2}{\alpha_{prior}}$.

[23]The gamma distribution can also be parameterized using shape k and scale $\theta = \beta^{-1}$.

Figure 8: Examples of Belief Distributions of Unknown Variance
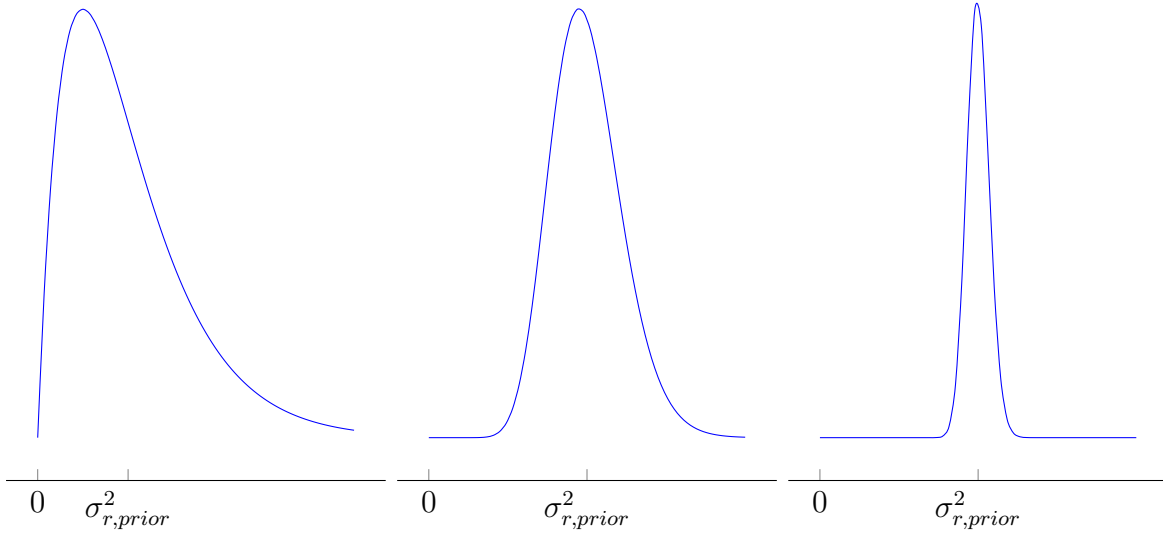
This figure illustrates three examples of gamma distributions that represent beliefs of the unknown parameter, σ_r^2, the variance of the return distribution. The gamma distribution is used since all variances are non-negative. All three represent the same belief: $\sigma_{r,prior}^2$. From left to right, the tightness or conviction in beliefs is increasing (i.e., the variance of the gamma distribution is decreasing). Note that in our analysis, $\sigma_{r,prior}^2$ drives the width of the confidence interval. The variance of the belief determines the strength of the prior on the confidence interval.

updated value for β incorporates the additional squared error from this new observation.[24]

6.1.3. The Persistence of Miscalibration

The model suggests why we may observe the persistence of severe miscalibration. First, in the model, there are mechanically diminishing returns to learning, because as the forecaster observes new information, they hold their beliefs with stronger conviction. Each additional observation, regardless of its relation to the prior belief, mechanically increases the denominator of the updating rule that defines $\sigma_{r,posterior}^2$. As a result, the newly observed return, which enters the numerator of the updating rule and interacts with the prior belief, is weighted less for each new observation.

Second, as shown in Table 2, miscalibrated CFOs believe the variance is much smaller than the historical return data would suggest. This means that their belief of the variance, $\sigma_{r,prior}$, is smaller than it should be. Let $\beta_{historical}$ denote the historical sum of squared differences, and $\alpha_{historical}$ denote the number of observed returns. Then the correct belief of

---

[24]The halves in both expressions are technical artifacts from the density of the normal distribution and should not be over-interpreted. The parameter α is used as an exponent and square-halved. With β, the half arises from the inner summation in the probability density function of the normal distribution.

the variance, $\sigma^2_{r,historical}$, is given by

$$\sigma^2_{r,historical} = \frac{\beta_{historical}}{\alpha_{historical}},$$

but a miscalibrated CFO's belief is such that $\sigma^2_{r,prior} < \sigma^2_{r,historical}$. In our model, this means that either the CFO believes the sum of squared differences is less than the historically observed sum of squared differences, i.e., $\beta_{prior} < \beta_{historical}$, and/or that their belief was generated using more observations than in the historical distribution, i.e., $\alpha_{prior} > \alpha_{historical}$.

The updating expression shows that miscalibration directly impacts the rate of forecaster learning. The larger is α_{prior}, the larger is the denominator, implying less updating in response to newly observed information. This is related to the insight in Section 3.4 that in forming priors, a smaller belief of the variance implies that the belief of the variance is more strongly held, which intuitively implies less updating in response to new information. Further, the larger is α_{prior} or the smaller is β_{prior}, the further the forecaster begins from the truth, thus requiring more updating to become properly calibrated.

Altogether, the observed miscalibration implies that in addition to beginning with a prior belief far away from the correct belief, updating is dampened. As a result, miscalibration may persist, even with rational Bayesian learning.

6.2. A Model of Conditional Variance

Given the well-known heteroskedasticity of returns, we now consider a more realistic model of conditional variance. We continue to assume that forecasters believe returns are the sum of a constant mean, $\bar{r}$, and an unanticipated shock, ϵ_t:

$$r_t = \bar{r} + \epsilon_t.$$

Now, however, forecasters believe that volatility follows an ARCH(1) process and is thus is composed of two parts:

$$\epsilon_t = e_t \sigma_t$$
$$e_t \sim N(0,1)$$
$$\sigma_t = \omega + \gamma \epsilon^2_{t-1}.$$

With an ARCH(1) process, volatility in the current period is the sum of a constant $\omega > 0$ plus an autoregressive component related to volatility in the previous period. The autoregressive component is parameterized by $\gamma > 0$, which multiplies the squared unexpected return observed in the previous period, $\epsilon^2_{t-1} = (r_t - \bar{r})^2$. Using this model, it is straightforward to

derive that a forecaster predicts higher volatility in the next period relative to the current period, $\sigma_{t+1}^2 > \sigma_t^2$, whenever

$$|r_t - \bar{r}| > |r_{t-1} - \bar{r}|.$$

This simple rule states that when the absolute forecast error in this period is larger than it was in the previous period, then conditional volatility in the next period will likewise increase. This is hardwired due to the (positive) autoregressive component of the process. Crucial to deriving this simple rule is the assumption that the mean is constant over time, but even without this simplification, a similar rule can be derived.

We introduce Bayesian uncertainty by weakening the strong assumption that forecasters can fully characterize the underlying volatility process. Instead, we now assume that forecasters are uncertain about the magnitude of autoregressive parameter, γ. To further simplify the exposition, we set $\omega = 0$ and therefore $\sigma_t^2 = \gamma \epsilon_{t-1}^2$. The forecaster's task is to again generate a prediction for next period's volatility, $\sigma_{t+1}^2 = \gamma \epsilon_t^2$. To do this, the forecaster calculates the squared forecast error ϵ_t^2 and a belief for the autoregressive parameter γ.

Once r_t is observed, the squared error is straightforward to construct. As before, the forecaster uses the newly observed return to update her belief about the unknown parameter. In particular, at the start of period t and before observing r_t, we assume that her prior beliefs, γ_{prior}^{-1}, are distributed according to a gamma distribution, $\Gamma(\cdot)$, since the autoregressive parameter must be strictly positive.

Given that $\gamma_{prior}^{-1} \sim \Gamma(\alpha_{prior}, \beta_{prior})$, the scale property of the gamma distribution yields that $\sigma_t^{-2} = \gamma_{prior}^{-1} \epsilon_{t-1}^{-2} \sim \Gamma(\alpha_{prior}, \beta_{prior} \epsilon_{t-1}^2)$. The derivation then proceeds exactly as above, yielding:

$$\alpha_{posterior} = \alpha_{prior} + \frac{1}{2}$$

$$\beta_{posterior} = \beta_{prior} \epsilon_{t-1}^2 + \frac{1}{2}(r_t - \bar{r})^2.$$

The resulting posterior distribution is also a gamma distribution with the updated parameters $\alpha_{posterior}$ and $\beta_{posterior}$. Using the posterior distribution, the forecaster's updated belief for the autoregressive parameter,

$$\gamma_{posterior} = \frac{\beta_{posterior} \epsilon_{t-1}^2}{\alpha_{posterior}},$$

is used to generate the forecast for next period volatility,

$$\sigma_{t+1}^2 = \gamma_{posterior} \epsilon_t^2.$$

The observed return, r_t, is used twice: once in the Bayesian-updated belief of $\gamma_{posterior}$ and once in defining ϵ_t^2. To disentangle the Bayesian learning from the autoregressive updating, we focus solely on changes in the belief of the autoregressive parameter. In particular, we can derive that:

$$\gamma_{posterior} > \gamma_{prior} \iff (r_t - \bar{r})^2 > \sigma_t^2.$$

The condition on the right hand side can be rewritten as $r_t \notin [\bar{r} - \sigma_t, \bar{r} + \sigma_t]$. If the observed return falls outside a one standard deviation interval around the known mean, $\bar{r}$, then the forecaster updates her belief of the autoregressive parameter, γ, upwards.

All else equal, increasing the autoregressive parameter increases the forecast of next period volatility. Similarly, if the observed return falls within this interval, the forecaster updates her belief downward, resulting in a decrease in the variance forecast. Overall, however, because the new belief of volatility also incorporates the new squared error, ϵ_t^2, understanding the effects of parameter learning requires controlling for changes in volatility that are predicted by the autoregressive component of the model.

6.3. Using the Bayesian Model to Interpret the Empirical Results

6.3.1. Are CFOs Bayesian Learners?

Our baseline results on updating are consistent with forecasters' using Bayes' rule to update their beliefs of the variance. When forecasters hit the CI, the new interval narrows, and when forecasters miss the CI, the new interval widens. We call this learning.[25] Stepping outside the simple model, we further find evidence that forecasters adjust their lower and upper intervals separately. Forecasters update the upper interval more when they miss on the upside, and update the lower interval more when they miss on the downside. CFOs may hold separate beliefs of positive and negative returns and update both intervals in response to news.

Of course, forecasters do not literally use Bayes' rule to construct their beliefs, but our evidence suggests that forecasters' beliefs evolve by combining prior beliefs with new information, as opposed to, for example, pure extrapolation (Barberis et al., 2015, Hirshleifer et al., 2015, Jin and Sui, 2019). Therefore, our evidence supports the broad class of models built on Bayes' rule that study under- and over-reaction to news (for example, Gervais and Odean (2001), Lewellen and Shanken (2002), Martin and Papadimitriou (2020), Bordalo et al. (2020)), without selecting a particular method of learning. Bordalo et al. (2020) is a closely related paper that studies how agents form beliefs of stock returns when they have

---

[25]Appendix B addresses the robustness of our results. We find evidence of learning when we explicitly control for regression to the mean, and we rule out the possibility that forecasters' update their beliefs using the simple ARCH model presented in Section 6.2.

diagnostic expectations. Similar to our results, they find that forecasters under-react to news about one-year-ahead stock returns (see Section III.B of the Internet Appendix). In their model, under-reaction is due to diagnostic beliefs, a separate mechanism than overconfidence. Our model and empirical findings suggest a complimentary explanation that CFOs strongly hold their initial beliefs, dampening updating and inducing persistent miscalibration.

6.3.2. Why Does Miscalibration Persist?

Our main finding is that despite learning, CFOs' confidence intervals remain far too tight. The magnitudes of the changes in CI width shed light on why: although forecasters are indeed updating their intervals in the appropriate direction when they miss, the size of the change, though economically significant, is not enough to attain proper calibration. In the model, there are two reasons why this may occur. First, with each subsequent updating, the conviction in the belief increases, dampening further learning. Second, the greater is a forecaster's initial miscalibration, the less is their learning from new information.

The evidence on path dependent learning is also in line with the narrative of persistent miscalibration despite (Bayesian) learning. To see this, consider a miss-hit forecaster, who initially observes a realized return and that it falls outside of their interval. According to the model, this has two consequences. First, as the forecaster forms their new belief, the additional observed return induces a stronger conviction in the new belief. Second, since the return falls outside the interval, the subsequent interval is wider. In the next period, the forecaster observes another return and that it falls inside their new interval. The key is that since the belief is now held more strongly than in the previous period, there is relatively less updating than in the previous period. The forecaster updates their beliefs and the new confidence interval is narrower, but the net effect over the two updates is that the interval widens. Similarly, consider a hit-miss forecaster. After initially hitting the interval, there are again two consequences. First, they form their new belief and it is held with a stronger conviction. Second, in contrast to above, the subsequent interval is narrower because the the return fell inside the interval. In the next period, the observed return falls outside the interval, inducing a widening. As above, since the belief is held more strongly than in the previous period, the updating is dampened, and the net effect over the two updates is that the interval narrows.

7. The Dynamics of Individual Forecasters

In this section, we analyze the behavior of several forecasters for whom we observe a long series of forecasts. These forecasters are not chosen at random. First, we highlight two forecasters whose behavior is consistent with our model, then discuss one forecaster whose

behavior is less consistent with our models and discuss the implications. For each survey date, the grey area represents the previously forecasted CI from one year ago, the red marker is the realized return, and the black interval is the newly forecast CI for the one-year-ahead return.

7.1. Forecaster #1363

Figure 9: Evolution of CIs for Forecaster #1363
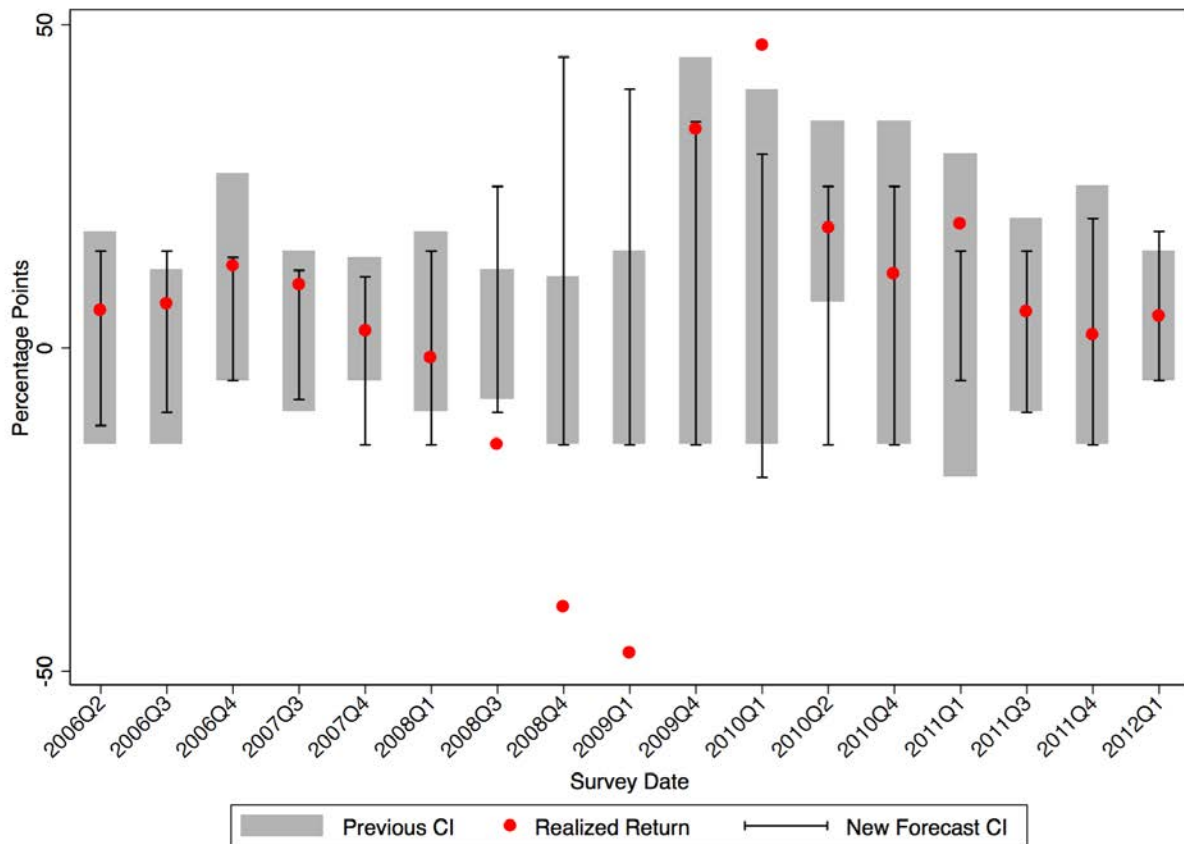
Evolution of confidence intervals for Forecaster #1363. For each survey date, the gray area is the one-year-ago forecast of the confidence interval. The red circle is the realized return at each date, and the black interval is the new forecast CI for the one-year-ahead return. The average CI width for all responses is 34.1 pp.

Figure 9 shows the evolution of CI widths for forecaster #1363. We have 17 observations for which we can observe the previously forecasted CI, the realized return, and the new forecast. This forecaster's behavior is largely consistent with the predictions of our Bayesian model. In the first four observations of Figure 9, the forecaster hits the CI and subsequently narrows their confidence interval. In the fifth and sixth observations, corresponding to

37

2007Q4 and 2008Q1, the forecaster hits the CI and widens their new interval which is inconsistent with learning. In both cases, the forecaster lowers the upper bound of the interval, but also lowers the lower bound of the interval by a large enough amount such that the net effect is to widen the CI.

The forecaster misses the CI for the next three observations and, in each case, widens the confidence interval. It is interesting that in 2008Q4 and 2009Q1, despite missing the CI low by more than 20%, both times the forecaster increases the upper bound of the CI and leaves the lower bound unchanged. This is further evidence that the forecaster remembers their previous interval between surveys. In 2010Q2, the forecaster hits the CI but widens the interval. For the next four observations, the forecaster hits the CI and subsequently narrows the interval. In the last observation, the forecaster hits the CI and very slightly widens the new interval.

7.2. Forecaster #2265

Figure 10 illustrates the evolution of Forecaster #2265's 17 confidence intervals. We again find that the forecaster's behavior is largely consistent with Bayesian learning, especially in the later surveys. Of the first five observations, three are inconsistent with the predictions of our model. In 2010Q1 and 2011Q1, the forecaster misses and narrows, while in 2011Q2, the forecaster hits and widens. The remaining 12 observations all display behavior consistent with our model: missing widens the CI, while hitting narrows it.

We note that the forecaster's later behavior is more consistent with our model. This may imply that with more experience in making forecasts, Forecaster #2265 becomes a "better" Bayesian learner.

7.3. Forecaster #126

Forecaster #126 is an example of a CFO who does not appear to be a Bayesian learner as described by our model. In the first six observations in Figure 11, the forecaster's behavior is consistent with our model only once. In the first observation, for 2006Q4, the forecaster misses high and though they move the interval upward, the overall width is smaller. In 2008Q1, the forecaster hits the CI and narrows, as predicted by the model. In the next four observations, between 2010Q3 to 2011Q2, the forecaster misses all four times but either keeps the CI roughly the same or narrows it.

The forecaster then hits the CI three times from 2011Q3 to to 2012Q1 but, each time, subsequently widens the CI. For the next five observations from 2012Q3 to 2013Q4 (excluding 2013Q3, for which we have no response), the forecaster misses the CI each time but narrows their next CI. In each instance, the forecaster misses high and, as expected, raises their lower
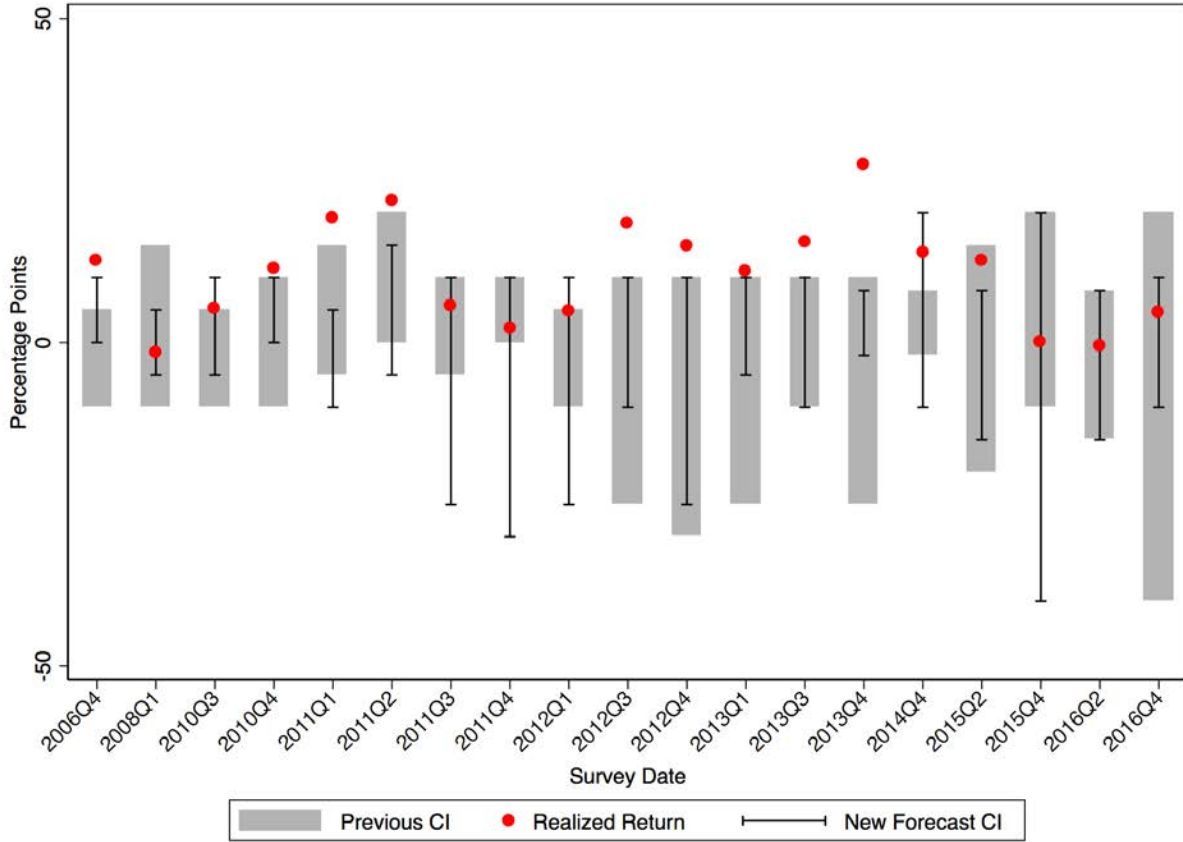
Figure 10: Evolution of CIs for Forecaster #2265
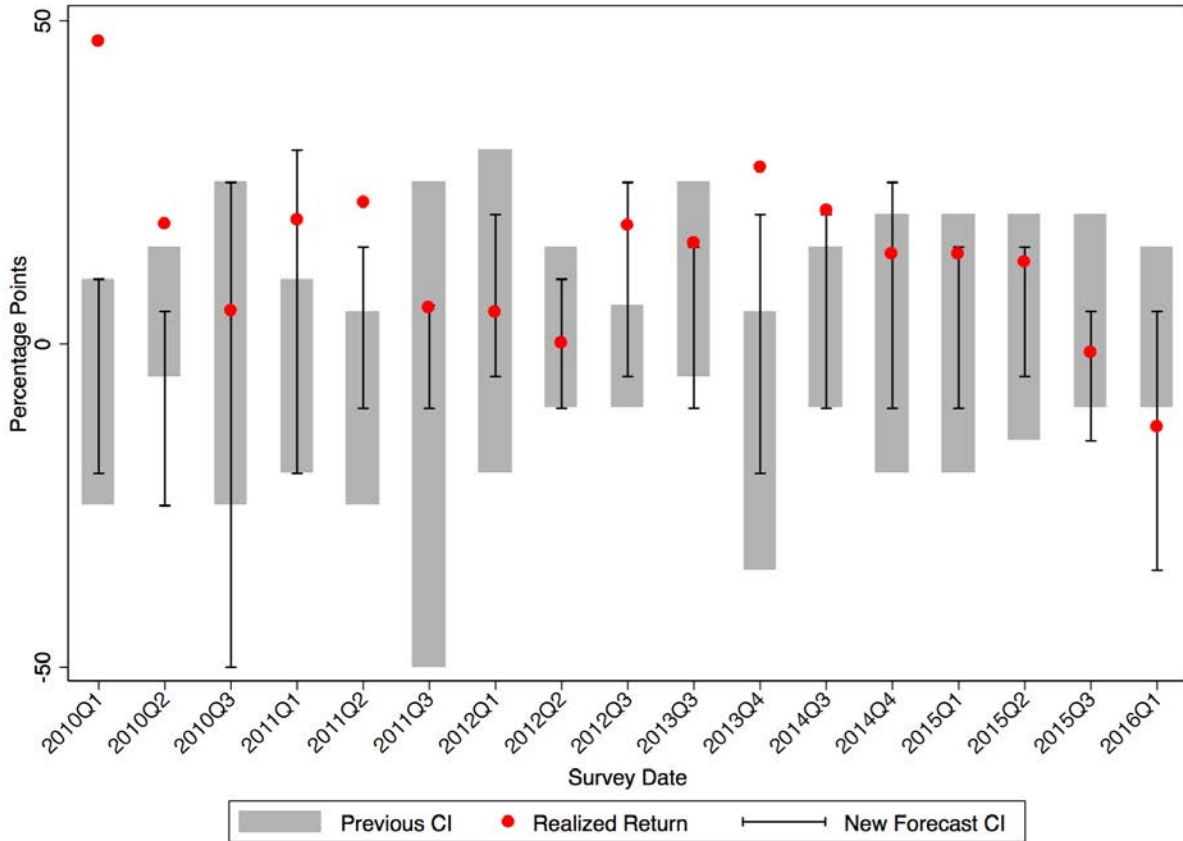
Evolution of confidence intervals for Forecaster #2265. For each survey date, the gray area is the one-year-ago forecast of the confidence interval. The red circle is the realized return at each date, and the black interval is the new forecast CI for the one-year-ahead return. The average CI width for all responses is 31.5 pp.

bound. However, the forecaster keeps almost the exact same upper bound in each case, thus narrowing the total interval length.

The forecaster's behavior is consistent with our model in four of the final five observations. Overall, however, the forecaster's behavior is not in line with the predictions of our Bayesian model. At the same time, the evidence suggests some form of systematic updating on the forecaster's part. Though the forecaster is not a Bayesian learner as described by our model, we cannot rule out that they are using some other model to update their beliefs.

Figure 11: Evolution of CIs for Forecaster #126

Evolution of confidence intervals for Forecaster #126. For each survey date, the gray area is the one-year-ago forecast of the confidence interval. The red circle is the realized return at each date, and the black interval is the new forecast CI for the one-year-ahead return. The average CI width for all responses is 23.5 pp.

8. Conclusions

Senior financial executives' beliefs of the mean of the distribution of S&P 500 returns are accurate, but they appear to systematically underestimate the variance of this distribution. We refer to this as miscalibration, a type of overconfidence. In the introduction, we noted that four CFOs provided more than 40 forecasts. Figure 12 shows the CFO with 44 responses, the largest number of responses in our sample. In this case, the initial confidence interval was only 5%, indicating very poor calibration. Over the next 17 forecasts, the CFO learned and increased their cumulative average CI width to 13%. This increase of 160% is economically significant, but still not enough to attain the proper 40.6% width consistent with historical returns. By the 44th forecast, the average confidence internal is still about 13%. Interestingly, this CFO has the widest final confidence interval among the four CFOs

40

that provided more than 40 forecasts.

Figure 12: Evolution of CFO's Belief of CI Width
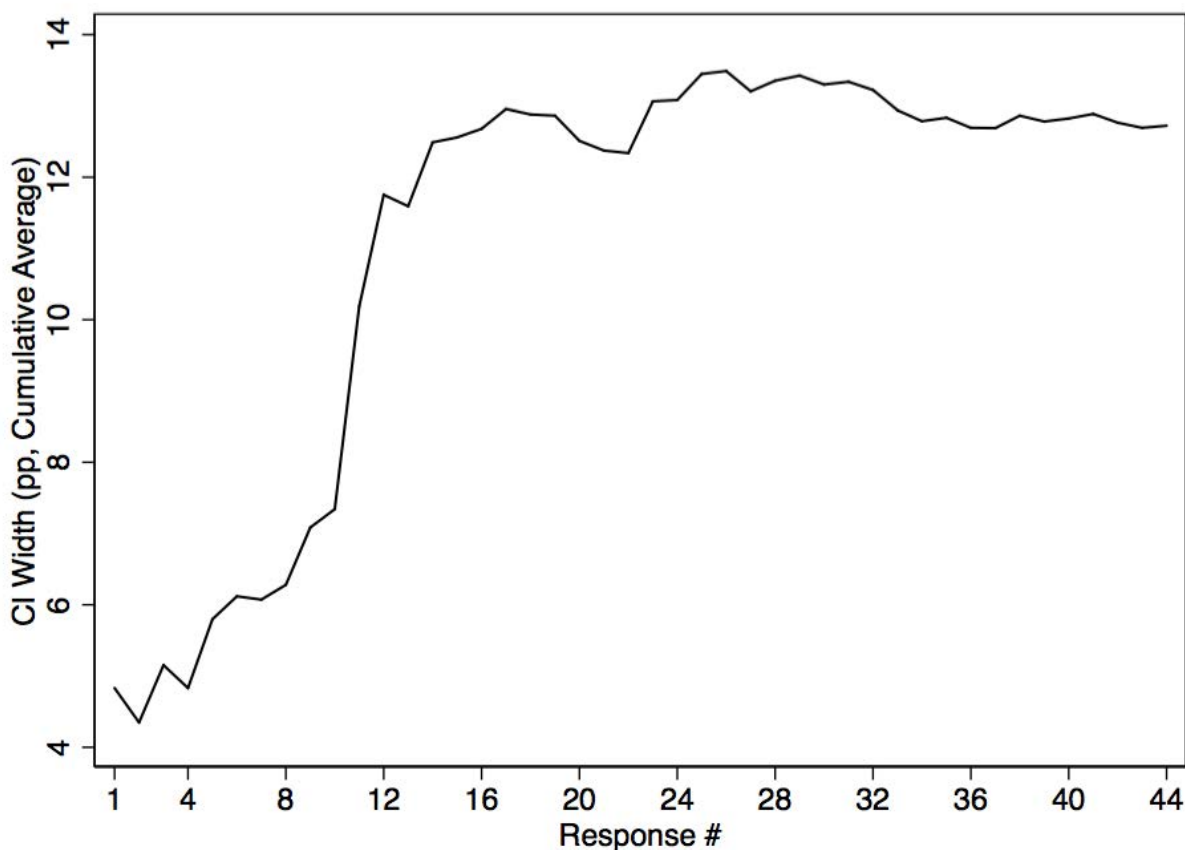
This figure plots the cumulative average CI width of the CFO with the largest number of responses in our sample (44 responses).

These results may be consistent with what Kahneman (2011, ch. 11) calls an "adjust-and-anchor heuristic." CFOs move from the anchor (in our context, the belief of the variance) with new information but "the adjustment typically ends prematurely" because they are "no longer certain than they should move further." Indeed, as David Viniar, then-CFO of Goldman Sachs, famously remarked in 2007, "The lesson you always learn is that your definition of extreme is not extreme enough."[26] Even after more than 40 forecasts, none of these four CFOs appear to have learned the definition of extreme, and the other three CFOs with long histories are even more miscalibrated than the one featured in this subsection.

Our research provides a unique analysis of how senior executives learn through time. We find that CFOs appear to update their beliefs of volatility in a manner consistent with

---

[26]See https://www.nytimes.com/2007/08/13/business/13cnd-goldman.html.

Bayesian learning. When the realized return falls outside of their forecasted confidence interval, they increase their belief of the variance. CFOs also update their beliefs asymmetrically. For example, after missing the interval on the upside, CFOs adjust the upper bound of their intervals more so than the lower bound.

Despite this learning, miscalibration persists. While the evidence shows that the updating is large and economically significant, it is not large enough to attain proper calibration. Our Bayesian model demonstrates that this sluggish updating may be connected to the strong conviction with which CFOs hold their beliefs. Further, the model predicts that miscalibration implies a strong conviction in beliefs. As a result, forecasters who are the most miscalibrated are also those who update the least in response to new information, which generates a cycle of persistent miscalibration. This is what we find in the data.

The framework that we have provided might be helpful, but it is not enough to completely resolve the puzzle of CFO miscalibration. We observe that CFOs learn and argue that the slow updating is consistent with very strong prior beliefs about the variance process. CFOs who miss the interval update on average by approximately 7.5 pp, but would need to update by an additional 23.1 pp to attain proper calibration. With enough time, our framework implies the CFOs should obtain proper calibration. However, even the forecasters with the longest track records do not converge.

References

J. S. Abarbanell and V. L. Bernard. Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance*, 47 (3):1181–1207, July 1992.

A. Ali, A. Klein, and J. Rosenfeld. Analysts' use of information about permanent and transitory earnings components in forecasting annual eps. *The Accounting Review*, 67: 183–198, 1992.

M. Alpert and H. Raiffa. *Judgment Under Uncertainty: Heuristics and Biases*, chapter A progress report on the training of probability assessors. Cambridge University Press, 1982.

B. M. Barber and T. Odean. The behavior of individual investors. In G. M. Constantinides, M. Harris, and R. M. Stulz, editors, *Handbook of the Economics of Finance*, volume 2, chapter 22, pages 1533–1570. Elsevier, 2013.

N. Barberis, A. Shleifer, and R. Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49:307–343, 1998.

N. Barberis, R. Greenwood, L. Jin, and A. Shleifer. X-capm: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115(1):1–24, 2015.

A. G. Barnett, J. C. van der Pols, and A. J. Dobson. Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1):215–220, February 2005.

J. M. Barrero. The micro and macro of managerial beliefs. Working paper, 2020.

R. B. Barsky and J. B. D. Long. Why does the stock market fluctuate? *Quarterly Journal of Economics*, 108(2):291–311, May 1993.

I. Ben-David, J. R. Graham, and C. R. Harvey. Managerial miscalibration. *Quarterly Journal of Economics*, 128(4):1547–1584, November 2013.

D. J. Benjamin. Errors in probabilistic reasoning and judgement biases. In B. D. Bernheim, S. DellaVigna, and D. Laibson, editors, *Handbook of Behavioral Economics: Applications and Foundations*, volume 2, chapter 2, pages 69–186. Elsevier, 2019.

N. Bloom, S. J. Davis, L. Foster, B. Lucking, S. Ohlmacher, and I. Saporta-Eksten. Business-level expectations and uncertainty. Working paper, December 2017.

P. Bordalo, N. Gennaioli, R. L. Porta, and A. Shleifer. Diagnostic expectations and stock returns. *Journal of Finance*, 2020.

J.-P. Bouchaud, P. Krüger, A. Landier, and D. Thesmar. Sticky expectations and the profitability anomaly. *Journal of Finance*, 74(2):639–674, April 2019.

D. Cesarini, M. Johannesson, P. Lichtenstein, and B. Wallace. Heritability of overconfidence. *Journal of the European Economic Association*, 7(2-3):617–627, Apr.-May 2009.

P. Collin-Dufresne, M. Johannes, and L. A. Lochstoer. Parameter learning in general equilibrium: The asset pricing implications. *American Economic Review*, 106(3):664–698, 2016.

A. Coutts. Good news and bad news are still news: experimental evidence on belief updating. *Experimental Ecomomics*, pages 1–27, 2018.

H. C. A. Dale. Weighing evidence: an attempt to assess the efficiency of the human operator. *Ergonomics*, 11(3):215–230, 1968.

W. Edwards. Conservatism in human information processing. In B. Kleinmuntz, editor, *Formal Representation of Human Judgment*, pages 17–52. Wiley, 1968.

W. Edwards and L. Phillips. Man as transducer for probabilities in bayesian command control systems. In M. Shelly and G. Bryan, editors, *Human Judgements and Optimality*, pages 360–401. Wiley, 1964.

L. Epstein and M. Schneider. Learning under ambiguity. *Review of Economic Studies*, 74(4):1275–1303, 2007.

L. Epstein and M. Schneider. Ambiguity, information quality and asset pricing. *Journal of Finance*, 63(1):197–228, 2008.

L. G. Epstein, J. Noor, and A. Sandroni. Non-bayesian updating: A theoretical framework. *Theoretical Economics*, pages 193–229, 2008.

E. Fermand, C. M. Kuhnen, G. Li, and I. Ben-David. Expectations uncertainty and household economic behavior. Working paper, 2019.

A. Fuster, D. Laibson, and B. Mendel. Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives*, 24(4):67–84, 2010.

A. Fuster, B. Hebert, and D. Laibson. Natural expectations, macroeconomic dynamics, and asset pricing. *NBER Macroeconomics Annual*, 26(1), 2011.

N. Gennaioli and A. Shleifer. What comes to mind. *Quarterly Journal of Economics*, 125(4):1399–1433, 2010.

N. Gennaioli, Y. Ma, and A. Shleifer. Expectations and investment. In M. Eichenbaum and J. Parker, editors, *NBER Macroeconomics Annual*, volume 30, pages 379 – 431. 2016.

S. Gervais and T. Odean. Learning to be overconfident. *Review of Financial Studies*, 14(1): 1–27, Spring 2001.

S. Giglio, M. Maggiori, J. Stroebel, and S. Utkus. Five facts about beliefs and portfolios. Working paper, 2019.

R. Greenwood and A. Shleifer. Expectations of returns and expected returns. *Review of Financial Studies*, 27(3):714–746, 2014.

H. Guo and J. A. Wachter. "Superstitious" investors. Working paper, 2019.

T. Henckel, G. D. Menzies, P. G. Moffatt, and D. J. Zizzo. Belief adjustment: A double hurdle model and experimental evidence. Working paper, 2018.

D. Hirshleifer, J. Li, and J. Yu. Asset prices in production economies with extrapolative expectations. *Journal of Monetary Economics*, 76:87–106, 2015.

L. J. Jin and P. Sui. Asset pricing with return extrapolation. Working paper, 2019.

D. Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

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

T. Kuchler and B. Zafar. Personal experiences and expectations about aggregate outcomes. *FRB of NY Staff Reports*, November 2017.

C. M. Kuhnen. Asymmetric learning from financial information. *Journal of Finance*, 70(5): 2029–2062, October 2014.

J. Lewellen and J. Shanken. Learning, asset pricing tests, and market efficiency. *Journal of Finance*, 57:1113–1145, 2002.

U. Malmendier and G. Tate. Behavioral ceos: The role of managerial overconfidence. *Journal of Economic Perspectives*, 29(4):37–60, Fall 2015.

U. Malmendier, G. Tate, and J. Yan. Overconfidence and early life experiences: The effect of managerial traits on corporate financial policies. *Journal of Finance*, 66(5):1687–1733, Nov. 2011.

K. Marray, N. Krishna, and J. Tang. How do expectations affect learning about fundamentals? some experimental evidence. Working paper, 2020.

I. Martin and D. Papadimitriou. Sentiment and speculation in a market with heterogeneous beliefs. Working paper, 2020.

M. Meeuwis, J. A. Parker, A. Schoar, and D. I. Simester. Belief disagreement and portfolio choice. Working paper, 2018.

M. M. Möbius, M. Niederle, P. Niehaus, and T. S. Rosenblat. Managing self-confidence: Theory and experimental evidence. Working paper, 2014.

A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, 3 edition, 1974.

D. A. Moore and D. Schatz. The three faces of overconfidence. *Social and Personality Psychology Compass*, 11(8), August 2017.

D. A. Moore, E. R. Tenney, and U. Haran. Overprecision in judgment. In G. Keren and G. Wu, editors, *Handbook of Judgement and Decision Making*. Oxford: Wiley Blackwell, 2016.

A. H. Murphy and R. L. Winkler. Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *National Weather Digest*, pages 2–9, 1977.

C. R. Peterson and L. R. Beach. Man as an intuitive statistician. *Psychological Bulletin*, 68 (1):29–46, 1967.

C. R. Peterson and W. M. Ducharme. A primacy effect in subjective probability revision. *Journal of Experimental Psychology*, 73(1):61–65, 1967.

L. Phillips and W. Edwards. Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72(3):364–354, 1966.

T. B. Roby. Belief states and sequential evidence. *Journal of Experimental Psychology*, 75 (2):236–245, 1967.

R. J. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71(71):421–36, June 1981.

P. Slovic and S. Lichtenstein. Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6(6):649–744, November 1971.

J. B. Soll and J. Klayman. Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):299–314, March 2004.

T. Strzaelcki. Axiomatic foundations of multiplier preferences. *Econometrica*, 79(1):47–73, January 2011.

J. D. Swart and R. Tonkens. The influence of order of presentation and characteristics of the datagenerator on opinion revision. *Acta Psychologica*, 41(2-3):101–117, April 1977.

R. Thaler. *Quasi Rational Economics*. Russell Sage Foundation, 1994.

A. Timmermann. How learning in financial markets generates excess volatility and predictability in stock prices. *Quarterly Journal of Economics*, 108(4):1135–1145, 1993.

P. Veronesi. Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *Review of Financial Studies*, 12(5):975–1007, 1999.

Appendix A. Bayesian vs. Knightian Uncertainty

This appendix briefly describes the differences, in our context, between Bayesian and Knightian uncertainty. Consider a CFO who knows that the return distribution is normally distributed and knows the variance, but is uncertain about the mean of the distribution.

A forecaster with Bayesian uncertainty forms a distribution of beliefs about the unknown parameter. For example, assume that her distribution of beliefs about the mean is normally distributed around 10%, with some variance that indicates the conviction of her belief. In this sense, she conceptualizes a "distribution of distributions" with respect to the return distribution. On the other hand, a forecaster with Knightian uncertainty may believe that the mean of the return distribution is either 8% or 12%. Knightian uncertainty is operationalized by specifying preferences that collapse the set of possible distributions into a single distribution.[27]

In this simple example, if the Knightian CFO were to treat the two possible distributions as two random variables and linearly combine them with equal weight, her beliefs would be equivalent to those of the Bayesian CFO. In general, agents with Knightian certainty combine their beliefs using some nonlinear rule, such as worst-case beliefs or multiplier preferences (see, for example, Strzaelcki, 2011).

In reality, Bayesian and Knightian uncertainty likely coexist. At times, there might be heightened Knightian uncertainty about the entire distribution of future returns. In what follows, we assume that forecasters have a good idea of the underlying distribution but are only Bayesian-uncertain about the exact parameters.

Appendix B. Robustness to Alternative Updating Rules

Appendix B.1. Increases or Decreases in Absolute Forecast Error

In this section, we assess whether the observed updating is consistent with the simple ARCH(1) model in Section 6.2 without Bayesian learning (i.e., with paramater certainty). We determine whether the absolute forecast error, $|r_t - \bar{r}_t|$, is higher for the survey at time t versus at time $t - 1$. If so, and the forecaster is learning, then the forecast for volatility, σ_{t+1}, should be higher than the previous forecast for volatility and the confidence interval

---

[27]We focus on Bayesian learning as a natural starting point but our analysis is also related to the broader literature on non-Bayesian learning (e.g., Epstein et al., 2008)). In particular, Epstein and Schneider (2007) develop a method for learning under Knightian uncertainty. Applying this method to asset prices, Epstein and Schneider (2008) show that ambiguity averse investors respond more to negative news than to positive news. In our context, a Bayesian agent is certain about the type of distribution of returns, but learns and updates her beliefs about the parameters governing the distribution. In contrast, a Knightian agent is unsure of the distribution of returns, and learns about the possible set of distributions.

Table A1: The Impact of Increases in Absolute Forecast Error
on Confidence Intervals

	Δ CI Width		
	(1)	(2)	(3)
Abs. Error Increased	1.65***	-0.80	0.49
	(0.64)	(0.86)	(0.59)
Unexpected Vol.			0.17***
			(0.03)
Forecaster Fixed Effects	Y	Y	Y
Time Fixed Effects	N	Y	N
Volatility Control	N	N	Y
Observations	1,875	1,875	1,875
R^2	0.006	0.098	0.030

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in CI width on indicator that activates when absolute forecast error increases between the realized return today and the realized return four quarters ago. Regressions include $1,875$ forecasts for which the four- and eight-quarter-ago forecast by the same respondent is available. The four-quarter-ago response is required the calculate the absolute forecast error for this period. The eight-quarter-ago forecast is required calculate the absolute forecast error from four quarters ago. Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

should widen. If the absolute forecast error is instead smaller, the confidence interval should narrow.

Table A1 presents estimation results from regressing the change in CI width on an indicator variable that activates when the absolute forecast error has increased. In the first column, without controlling for unexpected volatility, an increase in absolute forecast error leads to a significant widening the confidence interval by approximately 1.65 pp. Given that the average CI is 15.6% wide, this represents a modest 11% widening.

With time fixed effects, we no longer estimate a significant effect of the increase in absolute forecast error on changes in interval widths. We find similarly imprecise effects if we control only for unexpected volatility using either the GARCH or RV specifications. This is consistent with periods of high unexpected volatility yielding wider intervals regardless of absolute forecast error. These results are inconsistent with the simple model presented above, and we conclude that forecasters do not update their beliefs of volatility using a simple ARCH model.

Appendix B.2. Regression to the Mean

The key idea behind learning is that forecasters respond to new information. In its simplest form, learning may involve using observed returns to inform beliefs. In our context, forecasters learn by combining prior beliefs with newly observed returns to form posterior beliefs.

Alternatively, there are theories of the evolution of beliefs that do not require forecasters to respond to new information. One such theory that may be observationally consistent with our form of learning is regression to the mean (see Kahneman, 2011, chapter 17). This statistical phenomena manifests itself in many ways. Consider a simple example of the "Sports Illustrated Cover Jinx," defined by Wikipedia as "an urban legend that states that individuals or teams who appear on the cover of *Sports Illustrated* magazine will subsequently be jinxed (experience bad luck)." In most cases, however, the so-called jinx can easily be explained by regression to the mean. For an athlete to appear on the cover of *Sports Illustrated*, they must have performed extremely well, both by their own personal standards but also in comparison to the multitude of their competitors. It is reasonable to think that this athlete will be unable to outdo their own spectular performance in the subsequent year, and will thus regress back towards their regular performance. What appears to be a curse is merely the effect of being selected based on abnormally high performance, which, statistically, is unlikely to persist.

In our context, suppose our sample includes the extreme scenario of a CFO with a constant belief of the variance, i.e., she does not update her belief at any point in time. Despite the constant belief, however, it is reasonable to assume there may be some small randomness in the CFO's recall of her belief or in the way she fills out the survey. This randomness is zero on average but may induce the forecaster to report beliefs either smaller or larger than her true belief. These random shocks would induce variation in our data, despite the fact that her beliefs are constant. Further, if in one survey the reported belief of the variance is smaller than the true belief due to a negative random shock, then by regression to the mean, we should expect that in the next survey, the reported belief will be closer to the true belief and thus larger. Similarly, if the reported belief is larger than the true belief, then we would expect that in the next survey, the reported belief will be smaller.

As noted above, these changes are observationally equivalent to the evolution of beliefs we would expect from a Bayesian learner. However, it seems reasonable to believe that forecasters' beliefs evolve due to both learning and regression to the mean. To test for the presence of regression to the mean, we run the following regression, as in Barnett et al.

(2005):

$$\text{CI}_{it} = \alpha + \beta \cdot \text{Miss CI}_{it} + \xi \cdot (\text{CI}_{i,t-1} - \overline{\text{CI}}_{t-1}) + \gamma_i + \omega_t + u_{it}, \tag{B.1}$$

where CI_{it} is the CI width for forecaster i at time t, $\overline{\text{CI}}_t$ is the cross-sectional average CI width at time t, and Miss CI_{it} is an indicator that activates when the realized S&P 500 return at time t misses the CFO's interval produced at time $t-1$. The coefficient β measures the difference in CI width for a forecaster who missed the interval with their last prediction relative to one who hits it. Forecaster and survey-quarter fixed effects are represented by γ_i and ω_t, respectively.

The term $(\text{CI}_{i,t-1} - \overline{\text{CI}}_{t-1})$ controls for regression to the mean by measuring the distance between the cross-sectional average CI width and the individual forecaster's CI width. Since the average effect of randomness is zero, there should be no randomness present in the cross-sectional average.[28] Thus the larger is this difference, the more the forecaster's prediction was from the average forecast. We should thus expect a negative estimate of ξ. Indeed, the estimated value of ξ is -0.002, but the standard error is 0.017 and thus the estimate is not statistically different than zero.

The estimated value of β is 1.20 with standard error 0.392, which is significant at less than the 1% level. This estimate says that, controlling for regression to the mean, the average CI width of a forecaster who missed the interval four quarters ago is 1.20 pp wider than that of a forecaster who hit the interval. This estimate is consistent with our analysis above that forecasters who miss the interval widen their subsequent predictions.[29] Thus even in the presence of regression to the mean, we find evidence consistent with Bayesian learning.

Appendix C. Additional Results

Appendix C.1. Changes in Lower and Upper Bounds of the Confidence Interval

This section analyzes changes in the upper and lower bounds of the interval when forecasters hit or miss their CIs. Table A2 shows the results from regressing the changes in these bounds on an indicator that activates when the forecaster missed their CI. These results are directly comparable to those in Table 4.

---

[28]We perform a similar analysis using the distance between the forecaster's individual time-series average CI width and the current predicted width. The results are similar.

[29]We note that, on its own, this estimate does not necessarily imply that forecasters who missed the interval widened their subsequent intervals. For example, it may be the case that forecasters who miss narrow by 1.0 pp and forecasters who hit narrow by 2.2 pp. For this reason, the analysis in this section is complementary to the analysis above.

For example, recall that with only forecaster fixed effects, missing the interval widened the CI by approximately 5.2 pp. The estimate in the first column shows that most of this change is driven by an increase in the upper bound. Similarly, when controlling for time fixed effects, the majority of the total change is driven by increasing the upper bound. However, there is a large and significant effect of widening from lowering the lower bound. This effect is smaller and less precisely estimated when controlling only for volatility time effects.

Table A3 presents estimates from regressing the changes in the upper and lower bounds on indicators that activate when forecasters miss the interval high or low. Controlling only for forecaster fixed effects, missing high raises the upper bound by 5.0 pp and the lower bound by 1.1 pp. The estimates are similar when controlling for volatility time effects. Controlling for more general fixed time effects, however, the estimated change on the upper bound when missing high is larger while the change in the lower bound is smaller.

With only forecaster fixed effects, forecasters who miss low increase their upper bounds by 3.5 pp and decrease their lower bounds by 4.8 pp. Controlling for general time fixed effects, we find instead that missing low lowers both the upper and lower bounds. The estimates when controlling for only volatility time effects imply that the volatility controls explain most of the estimated impact of the time fixed effects on changes in the lower bounds. The estimated effects on the upper bounds is not significant at traditional levels, implying again that there are other time fixed effects driving the changes.

Appendix C.2. Changes in Upper and Lower Parts of the Confidence Interval

In the first column of Table A4, we see that with only forecaster fixed effects, forecasters who miss the CI widen the upper portion of their intervals by approximately 2.2 pp. In the second column, the lower portion of the interval widens by approximately 3.0 pp. Together, these sum to the total widening of 5.2 pp reported in the first column of Table 4.

With forecaster and time fixed effects, forecasters who miss the CI widen the upper portion of their intervals by 2.8 pp more than those who hit the CI, and widen the lower portion of their intervals by 4.7 pp more than those who hit the CI. In the final two columns with volatility controls instead of general time fixed controls, these estimates decrease to 1.9 and 2.8 pp, respectively, implying that the volatility controls are only one part of the fixed time effects.

Table A2: The Impact of Missing the Interval on Upper and Lower Bounds

	Δ UB	Δ LB	Δ UB	Δ LB	Δ UB	Δ LB
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator: Miss CI	4.57***	-0.64	4.98***	-2.54***	4.30***	-0.45
	(0.34)	(0.40)	(0.43)	(0.56)	(0.32)	(0.41)
Unexpected Vol.					0.09***	-0.06***
					(0.02)	(0.02)
Exp. Change in Vol.					0.06	-0.10
					(0.06)	(0.08)
Forecaster Fixed Effects	Y	Y	Y	Y	Y	Y
Time Fixed Effects	N	N	Y	Y	N	N
Volatility Controls	N	N	N	N	Y	Y
Observations	4,643	4,643	4,643	4,643	4,643	4,643
R^2	0.073	0.001	0.144	0.090	0.090	0.008

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in the upper (UB) and lower (UB) bounds of the confidence interval on an indicator that activates when the realized return missed the confidence interval. Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

Table A3: The Impact of Missing the Interval High or Missing Low on Upper and Lower Bounds

	Δ UB	Δ LB	Δ UB	Δ LB	Δ UB	Δ LB
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator: Miss CI High	5.03***	1.11**	8.57***	-0.11	5.57***	1.32***
	(0.33)	(0.44)	(0.53)	(0.71)	(0.33)	(0.44)
Indicator: Miss CI Low	3.50***	-4.75***	-1.35**	-6.81***	-0.30	-6.90***
	(0.50)	(0.51)	(0.59)	(0.78)	(0.45)	(0.60)
Unexpected Vol.					0.25***	0.16***
					(0.03)	(0.03)
Exp. Change in Vol.					-0.02	-0.22***
					(0.06)	(0.07)
Forecaster Fixed Effects	Y	Y	Y	Y	Y	Y
Time Fixed Effects	N	N	Y	Y	N	N
Volatility Controls	N	N	N	N	Y	Y
Observations	4,643	4,643	4,643	4,643	4,643	4,643
R^2	0.079	0.052	0.201	0.105	0.139	0.063

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regressions of the change in the upper (UB) and lower (UB) bounds of the confidence interval on indicators that activate when the forecaster missed the interval high or low. Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

Table A4: The Impact of Missing the Interval on Upper and Lower Confidence Intervals

	Δ UCI	Δ LCI	Δ UCI	Δ LCI	Δ UCI	Δ LCI
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator: Miss CI	2.17***	3.04***	2.84***	4.67***	1.97***	2.79***
	(0.25)	(0.39)	(0.30)	(0.49)	(0.24)	(0.39)
Unexpected Vol.					0.07***	0.08***
					(0.01)	(0.02)
Exp. Change in Vol.					0.06*	0.10*
					(0.04)	(0.06)
Forecaster Fixed Effects	Y	Y	Y	Y	Y	Y
Time Fixed Effects	N	N	Y	Y	N	N
Volatility Controls	N	N	N	N	Y	Y
Observations	4,643	4,643	4,643	4,643	4,643	4,643
R^2	0.032	0.025	0.097	0.085	0.054	0.040

Standard errors (in parentheses) clustered at the forecaster level. *, **, *** denote significance at the $0.10, 0.05$, and 0.01 levels under the assumption of a single test. Regression of the change in sizes of the upper and lower portions of the confidence interval on an indicator that activates when the realized return missed the confidence interval. The upper portion of the confidence interval (UCI) is the distance between the upper bound and the forecast. The lower portion of the confidence interval (LCI) is the distance between the forecast and the lower bound. Economic Significance is the estimated change in size divided into the average initial size. The average initial upper CI is 6.0% and the average initial lower CI is 9.6%. Unexpected volatility is the difference between realized volatility and RV-based forecast of volatility (see Section 2.4). Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

Appendix D. Selection

One might worry that forecasters select into our sample. For example, only the forecasters who are more amenable to learning (as defined in Section 5) select into the sample, biasing our results upwards. A more natural assumption is that poor forecasters drop out of the sample, biasing our main results on learning. We think the selection issue is not a first-order problem for two reasons.

First, the Duke-CFO survey has a large number of question and the stock market forecasts are a minor part of the survey. We believe it is unlikely that a CFO would stop replying to the survey because of their performance on a single question. Second, the likelihood of participating in the survey four quarters after a CFO provides an initial forecast is 31.4%. For those who ex post missed the interval with their initial forecast, probability decreases by a modest 3.5 pp.

Appendix E. Additional Background Material on Learning in Psychology and Economics

Benjamin (2019) provides a detailed overview of the vast literature in psychology that has been studying overconfidence and the updating of beliefs for many decades. Beginning in at least the early 1960s, psychologists began studying Bayesian updating and deviations from perfect rationality (Edwards and Phillips (1964) and Phillips and Edwards (1966)). Broadly speaking, the analysis undertaken in this paper is related to the idea of biased sampling-distribution beliefs (see, for example, Peterson and Beach (1967), Edwards (1968), and Slovic and Lichtenstein (1971)). This is the notion that agents behave as fully rational Bayesian updaters, but their belief distributions are inherently biased. In our context, updating of belief distributions appears to be consistent with Bayesian learning, but these belief distributions are distorted in that their variances are too small. Our finding of insufficient updating is consistent with idea of anchoring in Kahneman (2011, ch. 11). He describes an "adjust-and-anchor heuristic" where, in our context, a CFO has an initial belief for the uncertain variance of S&P500 returns that acts as an anchor. As the CFO observes more information, she slowly moves away from the anchor. Kahneman argues that "the adjustment typically ends prematurely, because people stop when they are no longer certain that they should move farther."

In contrast to our analysis of CFOs forecasting real-world S&P 500 returns, psychologists primarily study the results of experiments that they design and implement in a lab. Many studies find evidence that at the individual-level, agents receiving signals under-react and update very little or not at all (see, for example, Möbius et al. (2014), Coutts (2018), and

Henckel et al. (2018)). Several experiments find a "primacy effect," where given a sequence of signals, those observed earlier have a larger impact on beliefs than those observed later (Peterson and Ducharme (1967), Roby (1967), Dale (1968), and Swart and Tonkens (1977)). Consistent with this literature, we find that CFOs tend to update less as they gain more experience in forecasting returns (as measured by the number of responses to the Duke-CFO survey).

Our paper is also broadly related to the reenforcement learning literature (see, for example, Barber and Odean (2013)). As Thaler (1994) notes, we should expect more learning and better calibration from forecasters who receive frequent and timely feedback to their forecasts, such as weathermen, and our unique dataset allows us to examine the behavior of financial officers who have many opportunities to learn. Much of the literature on learning uses expectations formed via extrapolation (see, for example, Barsky and Long (1993), Fuster et al. (2011), Hirshleifer et al. (2015), Barberis et al. (2015), Jin and Sui (2019), and Guo and Wachter (2019)). Another popular strand of the literature focuses on expectations formed by using "news" or any newly observed information (see, for example, Timmermann (1993), Barberis et al. (1998), Veronesi (1999), Lewellen and Shanken (2002), and Collin-Dufresne et al. (2016)). Fuster et al. (2010) present a model of "natural expectations" that combines standard rational expectations and adaptive expectations. In a sense, Bayesian learning incorporates both an extrapolative and information-based component into the formation of expectations. In our case, the extrapolative component comes from the prior, and the news is the newly observed stock market return.