Managerial Recalibration: Do CFOs Learn?

Michael Boutros  
*Duke University, Durham, NC 27708 USA*

John R. Graham  
*Duke University, Durham, NC 27708 USA*  
*National Bureau of Economic Research, Cambridge, MA 02138 USA*

Campbell R. Harvey  
*Duke University, Durham, NC 27708 USA*  
*National Bureau of Economic Research, Cambridge, MA 02138 USA*

Itzhak Ben-David  
*The Ohio State University, Columbus, OH 43210 USA*  
*National Bureau of Economic Research, Cambridge, MA 02138 USA*

First Draft (Incomplete and Preliminary)  
Current Version: October 21, 2018

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 that provide multiple forecasts to evaluate how they adapt and recalibrate their forecasts in response to return realizations. We focus on the confidence intervals that each CFO provides in their forecast. A simple model of Bayesian learning suggests that confidence intervals should be impacted by return realizations. We find that the confidence intervals significantly widen when realizations fall outside CFOs’ ex ante confidence ranges and narrow when realizations fall within their ex ante intervals. While these results are consistent with CFOs learning, we find that CFOs are still badly miscalibrated.

*Keywords:* Learning, Information, Behavioral Economics, Volatility Forecasts, Market Forecasts, Behavioral Finance, Bayesian Updating.

---

1Send correspondence to: Campbell R. Harvey (cam.harvey@duke.edu). We are indebted to our colleague John Payne who suggested we pursue learning dynamics.
1. Introduction

It is now widely understood that decisions are impacted by the circumstances of the decision maker’s environment. Children of the Great Depression are likely to act differently than people that did not experience such hardship (Malmendier and Nagel, 2011). Further differences have been found between wealthy and non-wealthy investors (Vissing-Jorgensen, 2003) and between wealthy and poor individuals (Das et al., 2017). Less is known about the way people may update beliefs in reaction to immediate experiences, nor why overconfidence persists given that people might update.

The goal of our paper is to test whether Chief Financial Officers (CFOs) learn through time. Ben-David, Graham, and Harvey (2013) (hereafter BGH), in a multi-year survey of executives, show that CFOs are, on average, miscalibrated: the confidence intervals for their forecasts of S&P 500 returns are far too narrow. We augment their data in two ways: first, we update the data and second, and more importantly, we construct a panel of individual forecasters. Thousands of CFOs in our sample make multiple forecasts, allowing us to test whether these CFOs learn from their mistakes.

Our database consists of 14,800 predictions of the one-year return on the S&P 500 from more than 2,800 forecasters. Remarkably, one CFO has made 43 forecasts. The respondents to the survey also provide 80% confidence intervals for their forecasts. These intervals are gauges of expected return volatility as well as the CFO’s perceived precision. We focus on the interval given that it is very difficult to determine what the right S&P 500 point forecast should be. 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 framework is a simple model of Bayesian learning that focuses on the confidence intervals. While Bayes’ rule is well known and tested in many laboratory settings, there is no practical way to engage senior financial executives in a multi-year laboratory experiment. Instead, we rely on the results of quarterly survey that has been ongoing for 20 years and administered electronically over the last 12 years.

We track individual CFOs and measure how their confidence intervals change over time. All CFOs are trying to forecast the same item, next year’s return on the S&P 500, something that CFOs pay attention to given that companies are routinely evaluated relative to the market return and it is unlikely they can explain their firm’s stock performance without reference to overall market movements.

We build a model of learning in which forecasters use Bayes’ rule to update their beliefs of the return process given new observations. The model predicts that when a forecaster misses her confidence interval, she should widen her next forecast’s confidence intervals, and
that when she hits her confidence interval, she should narrow her next confidence interval. Further, the degree to which the forecaster narrows or widens her confidence interval depends on the degree to which she hit or missed.

We determine whether CFOs’ confidence intervals react to whether they hit or miss their prespecified intervals. Since return volatility is time-varying and persistent, we carefully consider whether CFOs are responding to feedback from their own forecasts or merely to changes in predicted and unexpected volatility. We find that, in general, CFOs are more likely to miss their ex ante interval when there are unexpected changes in volatility, and that high volatility generally leads to wider forecast intervals.

Controlling for unexpected changes in volatility, we find strong evidence that CFOs recalibrate their forecast confidence intervals given the feedback of realized market returns. Intervals increase when CFO forecasts miss their ex ante intervals and intervals decrease when CFOs hit their range. Interestingly, the intervals are not necessarily symmetric around the forecast. For example, we find that CFOs who undershoot tend to lower the downside relatively more than the upside.

Given the evidence in BGH of confidence intervals being too narrow, missing the confidence interval tends to reduce the degree of miscalibration. However, for those that hit the interval, the intervals become more miscalibrated as the (accurate) forecasts fuels additional overconfidence. Nevertheless, the recalibration after missing is much greater than the narrowing after hitting the interval.

Although we find strong evidence of recalibration and learning, the forecasters in our sample who learn remain badly miscalibrated. We reconcile these two facts by showing that CFOs widen their intervals by far too little. The average CI width for a forecast that missed the interval is 12.8%, while the 80% interval of realized returns from 1950 to 2018 is 40.6%. Forecasters who miss the CI indeed widen their intervals by an economically significant amount, 5 percentage points or almost 50%, but that is not nearly enough to attain proper calibration. This insufficient degree of updating explains why miscalibration persists over time.

Our paper is related to the broad literature on overprecision and, in particular, overconfidence, defined as “excessive faith that [the agent] knows the truth” (Moore and Schatz, 2017). Overconfidence is well documented in many settings, notably amongst CFOs in BGH but also in many other smaller laboratory settings (see, for example, Soll and Klayman (2004) and Moore et al. (2016)). This literature documents the presence of overconfidence but stops short of testing whether overconfident agents re-evaluate their confidence and adjust in the presence of making mistakes.

On the theoretical side, Gervais and Odean (2001) build a model of traders who learn
about their abilities over time through experience. These traders are overconfident despite realizations, attributing successes to their own ability and failures to bad luck, thus becoming more overconfident over time. Empirically, the model is evaluated by looking at cross-sectional moments of market data, and not by examining the empirical beliefs of traders. Our paper is also broadly related to the reinforcement learning literature, summarized in Barber and Odean (2013): “the simplest form of learning may be to repeat behaviors that previously coincided with pleasure and avoid those that coincided with pain.”

Our work is related to the growing literature on learning in finance and macroeconomics. Consistent with our results on asymmetric learning, Kuhnen (2014) finds that people in the “negative domain” of returns often form relatively more pessimistic beliefs about the investment options available to them. This negative bias is driven by agents responding more to outcomes in the negative versus positive domain. We construct a model of Bayesian learning, tapping the literature built around the well-known Bayes’ rule.

Much of the literature on learning uses expectations formed via extrapolation (see, for example, Barsky and Long (1993), Fuster et al. (2010), Hirshleifer et al. (2015), Barberis et al. (2015), Jin and Sui (2018), and Guo and Wachter (2018)). 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)). 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.

Finally, our paper is related to the growing literature on the beliefs held by firms and their executive officers. 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 successfully generate subjective distributions based on previous observations of data.

In addition to BGH, Greenwood and Shleifer (2014) and Gennaioli et al. (2016) use the same CFO survey as in our paper. Greenwood and Shleifer (2014) analyze a long time series of investor expectations and their relationship to expected returns in standard finance models. Gennaioli et al. (2016) look at a sample of firms and analyze the relationship between expectations of growth and investment. In contrast to the learning mechanism developed in this paper, they find that firms often use simple extrapolation in forming their next-period beliefs of growth. Similarly, Kuchler and Zafar (2017) find that U.S. households often
extrapolate local home price changes when surveyed about beliefs of national home price changes over the next year.

Our paper is organized as follows. The second section provides out of sample evidence of BGH’s findings by almost doubling their sample size. Our model of Bayesian learning is presented in the third section. The fourth section details our data and provides some summary statistics. The main evidence of recalibration and learning is presented in the fifth section. The sixth section analyses the performance of individual forecasters. Some concluding remarks are offered in the final section.

2. Revisiting Managerial Miscalibration

BGH study S&P500 return predictions made by senior financial executives, the majority of whom are CFOs, 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.

BGH report that CFOs hit their 80% confidence intervals only 33.8% percent of the time, providing striking evidence of miscalibration. Our updated data provide an opportunity for out-of-sample evidence. Interestingly, since 2011Q3, CFOs only hit their confidence intervals 22.8% of the time, suggesting they have become more miscalibrated on average. Though this difference of almost 10% is economically large, it is not statistically significant at the usual levels of significance.

Over the full sample, CFOs hit their 80% confidence intervals only 30.8% of the time. Figure 1 illustrates the percentage of responses in each survey that hit the forecast interval. We calculate annual S&P500 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 percentage points wide. Only 3.5% of responses have confidence intervals at least 40.6% wide. These results using the full sample are very similar to those in BGH. In their sample, the 80% confidence interval should be approximately 42.2% wide and only 3.4% of responses had intervals at least that wide. Hence, the results of BGH are validated out-of-sample, and, if anything, CFOs appear more miscalibrated than in the original study.

In the original BGH, CFOs hit their 80% confidence intervals over the first part of the sample, from 2001Q2 to 2011Q3, 36.3% of the time. This minor difference is due to a slightly modified algorithm for cleaning and merging data across surveys.
2.1. Miscalibration and Volatility

CFO confidence intervals are likely influenced by perceptions of market volatility. The comparisons made above compare the unconditional return distribution with CFOs’ conditional forecasts. 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&P500 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. 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 to forecasters at the time. Therefore, in studying the evidence surrounding learning and the evolution of volatility beliefs, we must be careful to account for both expected and unexpected changes.

3. Modelling Learning

In this section, we build a simple model of normally distributed returns with time-varying volatility. We assume that forecasters believe returns are the sum of a constant mean, $\bar{r}$, and an unanticipated shock, $\epsilon_t$:

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

Forecasters believe that volatility follows an ARCH(1) process and is thus is composed of two parts:

$$ \epsilon_t = \epsilon_t \sigma_t $$

$$ \epsilon_t \sim N(0, 1) $$

$$ \sigma_t = \omega + \gamma \epsilon_{t-1}^2. $$

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, $\epsilon_{t-1}^2 = (r_t - \bar{r})^2$. Using this model, it is straightforward to derive that a forecaster forecasts higher volatility in the next period relative to the current
The percentage of survey responses with the ex-post return falling within the ex-ante 80% confidence interval. 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 across all surveys and each dashed black line is the subsample average.

This simple rule states that when the absolute forecast error in this period is larger than it was in the previous period, then total 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.

3.1. Bayesian Learning Model

Assuming that forecasters can completely characterize the underlying volatility process is a very strong assumption. We now assume that although forecasters know the process, they are uncertain about the magnitude of autoregressive parameter, $\gamma$. 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, $\sigma_{t+1} = \gamma \epsilon_t^2$. To do this, the forecaster needs to calculate the squared forecast error $\epsilon_t^2$ and a belief for the autoregressive parameter $\gamma$.

Once $r_t$ is observed, the squared error is straightforward to construct. We build a model of Bayesian learning where observing $r_t$ also informs the forecaster’s beliefs about the unknown parameter’s value. At the beginning of each period, the forecaster has a prior belief of $\gamma$. Once the return is observed, she updates her beliefs and forms a posterior belief. For reasons that will become clear shortly, we focus on the belief of $\gamma^{-1}$.

In particular, at the start of period $t$ and before observing $r_t$, we assume that her prior beliefs, $\gamma^{-1}_{\text{prior}}$, are distributed according to a gamma distribution. In a slight abuse of notation, we use $\gamma^{-1}_{\text{prior}}$ to refer to the mean of the distribution, which implies the forecaster summarizes her distribution of beliefs using the mean. We are agnostic about the source of these prior beliefs. In reality, they may be derived from many years of observing market returns, or they may be formed after overhearing a rival forecaster’s beliefs. In our model, all that matters is that her beliefs are characterized by the gamma distribution, $\Gamma(\cdot)$.

We choose to model prior beliefs with a gamma distribution for several reasons. First, this distribution is physically plausible since it forces the variance to be strictly positive. Second, when this prior is combined with an observed return using Bayes’ rule, the posterior is also a gamma distribution. This property makes the gamma 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.

The gamma distribution is parameterized by $\alpha$, governing the shape, and $\beta$, which is called the inverse scale or rate parameter. The shape parameter essentially determines the mass surrounding the peak of the probability density while the scale parameter governs how “spread out” it is. 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 $\alpha$ can be interpreted as the strength of the prior. The larger is this parameter, the larger the mass surround the peak, and the stronger is the prior.

Given that $\gamma^{-1}_{\text{prior}} \sim \Gamma(\alpha_{\text{prior}}, \beta_{\text{prior}})$, the scale property of the gamma distribution yields that $\sigma_{t}^{-2} = \gamma^{-1}_{\text{prior}} \epsilon_{t-1}^{-2} \sim \Gamma(\alpha_{\text{prior}}, \beta_{\text{prior}} \epsilon_{t-1}^2)$. Thus the probability of the variance being some

---

3The gamma distribution can also be parameterized using shape $k$ and scale $\theta = \beta^{-1}$. 

value of $\sigma_t^{-2}$ is given by the pdf of the gamma distribution:

$$P_{\text{prior}}(\sigma_t^{-2}) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_t^{-2})^{\alpha_{\text{prior}} - 1} \exp \left(-\sigma_t^{-2} \beta_{\text{prior}} \epsilon_{t-1}^2\right) \propto (\sigma_t^{-2})^{\alpha_{\text{prior}} - 1} \exp \left(-\sigma_t^{-2} \beta_{\text{prior}} \epsilon_{t-1}^2\right).$$

This density summarizes the prior distribution of the beliefs.\(^4\) The agent knows that returns are normally distributed with mean $\bar{r}$. Given some belief of the variance, she can calculate the probability of observing $r_t$ as:

$$P(r_t | \sigma_t^2) \propto (\sigma_t^{-2})^{\frac{1}{2}} \exp \left(-\sigma_t^{-2} \frac{1}{2} (r_t - \bar{r})^2\right).$$

Both of these densities are more easily expressed using the inverse of the variance, which motivated our choice above. Combining the data density with the prior density yields the posterior density of the variance belief after observing $r_t$:

$$P_{\text{posterior}}(\sigma_t^{-2}) = P_{\text{prior}}(\sigma_t^{-2}) P(r_t | \sigma_t^2) \propto (\sigma_t^{-2})^{\alpha_{\text{posterior}} - 1} \exp \left(-\sigma_t^{-2} \beta_{\text{posterior}} \epsilon_{t-1}^2\right)$$

where

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

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

The resulting posterior distribution is also a gamma distribution with the updated parameters $\alpha_{\text{posterior}}$ and $\beta_{\text{posterior}}$. The new value for $\alpha$ reflects the additional scaling required from using one more observation to inform the belief of the variance. The updated value for $\beta$ incorporates the additional squared error from this new observation.\(^5\)

\(^4\)In Bayesian analysis, it is often easier to work with the probability kernel, which drops constant scalars from the density function. In what follows, we refer to this kernel as the probability density. See Murphy (2007) for a more detailed derivation.

\(^5\)The halves in both expressions are technical artifacts from the density of the normal distribution and should not be over-interpreted. For $\alpha$, this parameter is used as an exponent and square-halved. With $\beta$, the half arises from the inner summation in the pdf of the normal density.
Using the posterior distribution, the forecaster’s updated belief for variance is given by
\[
\frac{\beta_{\text{posterior}}}{\alpha_{\text{posterior}}} = \gamma_{\text{posterior}} \epsilon_{t-1}^2
\]
and therefore the updated belief for the unknown parameter is
\[
\gamma_{\text{posterior}} = \frac{\beta_{\text{posterior}} \epsilon_{t-1}^2}{\alpha_{\text{posterior}}}.
\]
The forecast for next period’s volatility is:
\[
\sigma_{t+1}^2 = \gamma_{\text{posterior}} \epsilon_t^2.
\]
The observed return \( r_t \) is used twice: once in the Bayesian-updated belief of \( \gamma_{\text{posterior}} \) and once in defining \( \epsilon_t^2 \). 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 under what conditions \( \gamma_{\text{posterior}} > \gamma_{\text{prior}} \). After some straightforward algebra, we find that
\[
\gamma_{\text{posterior}} > \gamma_{\text{prior}} \iff (r_t - \bar{r})^2 > \sigma_t^2.
\]
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 \( \gamma \) upwards.

All else equal, increasing \( \gamma \) 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 forecast. Overall, however, because the new belief of volatility also incorporates the new squared error, \( \epsilon_t^2 \), understanding the effects of parameter learning requires controlling for changes in volatility that are predicted by the autoregressive component of the model.

3.2. Summary of Model Implications

To summarize, agents using a simple GARCH model will forecast a higher variance in the next period relative to this period whenever the absolute forecast error from this period is larger than it was in the previous period:
\[
\sigma_{t+1} > \sigma_t \iff |r_t - \bar{r}| > |r_{t-1} - \bar{r}|.
\]

\(^6\text{See Appendix B for the full derivation.}\)
This result is not surprising and is driven solely by the autoregressive nature of the model. When we incorporate parameter uncertainty regarding the value of $\gamma$, then the observation of $r_t$ yields a higher belief of $\gamma$ when the return falls outside a one standard deviation interval around the mean:

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

Controlling for the expected change in variance, belief of an increased value for $\gamma$ yields a higher expected variance.

4. Data

We use a set of stock market predictions made by financial executives in the quarterly Duke-CFO survey. A detailed description is available in BGH. The primary survey question we are interested in 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 gives both a point estimate and an 80% confidence interval. In total, we have 14,800 responses for which we can identify the respondent. Figure 2 illustrates that we are able to construct a very large panel of data 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.

4.1. Width of Confidence Intervals

The first three columns of Table 1 describe the widths of confidence intervals for all responses in our sample. The mean width of CFO confidence intervals is 14.3%. As noted above, the distribution of annual S&P500 returns implies an 80% confidence interval that is 40.6% wide. The 75th percentile of forecasters have a CI width of 20.0%, which is still only half of the interval implied by the return data. CFOs’ confidence intervals are far too tight around their point estimates, implying overconfidence.

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 hist the interval has an interval twice as wide as the forecaster who misses, and this pattern is consistent at other percentiles of the distribution.
Figure 2: Overview of Panel Length

Figure shows the number of forecasts given by respondents in our sample. For example, approximately 950 forecasters have responded twice.

To study learning, we analyze pairs of responses from the same respondent such that we can observe the initial confidence interval, the realized return, and a new confidence interval four quarters ahead. This reduces the number of observations to 4,643. The last three columns of Table 1 summarize the lengths of confidence intervals for the new confidence interval in each pair of observations. Overall, the width of the CIs for the repeat forecasters is only slightly larger than the one time forecasters, 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&P500 returns of 40.6%.

The Bayesian learning model assumes that the forecaster’s beliefs are symmetrically
Table 1: Confidence Interval Width Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>All Forecasts</th>
<th>Repeat Forecasts</th>
<th>Single Forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit</td>
<td>Miss</td>
<td>All</td>
</tr>
<tr>
<td>Mean</td>
<td>20.7</td>
<td>11.7</td>
<td>14.3</td>
</tr>
<tr>
<td>Median</td>
<td>18.0</td>
<td>9.0</td>
<td>10.0</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>10.0</td>
<td>5.0</td>
<td>6.0</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>30.0</td>
<td>15.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Observations</td>
<td>4,308</td>
<td>10,492</td>
<td>14,800</td>
</tr>
</tbody>
</table>

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 initial confidence interval. One Time 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.

centered around their point forecast. We construct a simple measure of asymmetry as:

\[ A = \frac{r_U + r_L}{2} - r_P, \]

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. When \( A < 0 \), the interval is skewed to the left.

Figure 3 presents the distribution of this measure of asymmetry. The mean asymmetry is approximately -1.5 percentage points 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.

4.2. Volatility

Controlling for both unexpected volatility as well as expected changes in volatility is important for our analysis. As such, we need to construct a model of expected volatility. We use two models of volatility to generate forecasts. First, we estimate a standard GARCH(1,1) model using quarterly data on S&P500 returns. Secondly, we estimate a quarterly AR(1) process directly on the series of realized volatility (RV), measured using one year of daily S&P500 returns. In both cases, to match the horizon of our survey question, we generate a four-step-ahead forecast as our measure of expected volatility in one 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
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 one-percentile of observations.

between current period realized volatility and the forecast of current volatility from the previous period. The two series can be summarized as:

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

The two series of unexpected volatility have a correlation of 0.51. Unsurprisingly, we find that unexpected changes in volatility are negatively correlated with forecasting accuracy. For each survey, we measure total calibration or forecasting accuracy as the percentage of responses where the ex-post return fell within the ex-ante interval. We align each survey to the volatility shock one year later when the realized return is observed. The correlation between forecasting
accuracy and unexpected volatility is similar for both models of unexpected volatility: $-0.22$
for the GARCH model and $-0.29$ for the RV model.

5. Recalibration

Guided by our framework, we turn to the data in order to study how forecasters update
their beliefs of volatility. Relative to the assumptions made in our model, we use both time-
varying mean beliefs and the asymmetries provided by asking respondents for the lower and
upper bounds separately. Building on the results from the model with parameter certainty,
we determine whether the absolute forecast error, $r_t - \hat{r}_t$, is higher for the survey at time
$t$ versus at time $t - 1$. If so, then the forecast for volatility, $\sigma_{t+1}$, should be higher than
the previous forecast for volatility and the confidence interval should widen. If the absolute
forecast error is instead smaller, the confidence interval should narrow.

Our models are simplifications of the true process for volatility. In reality, there are
many other factors driving volatility, and these other factors are captured by the unexpected
level of volatility relative to the models’ predictions. We therefore control for unexpected
volatility in our regressions using the series generated by either the GARCH- or RV-based
models described above.

In our Bayesian learning model, we emphasize the role of learning about the autoregres-
sive parameter in forecasting volatility. In particular, we analyze the effect of new return
realizations on the belief of the autoregressive parameter and, in turn, how this updated
belief affects forecasts of volatility. As detailed in section 3, a newly observed return updates
total volatility in two ways. First, by updating the squared error term, and secondly, by
updating the belief of the unknown parameter. To focus on the latter effect only, we control
for the first effect by adding the expected change in volatility directly into our regressions.

We then determine whether the realized return falls within the forecaster’s confidence
interval. If within the interval, and now controlling for both the expected change in volatility
and any unexpected volatility, then the forecast for next period volatility should less than
before and the confidence interval should narrow. If instead the return falls outside the
confidence interval, the new confidence interval should widen.

We used a fixed-effects model to control for each individual forecaster. As detailed in
section 4.2, we construct the forecasts using both GARCH- and RV-based models. Using
the GARCH-based measures of volatility is consistent with our model but makes a very
strong assumption regarding how volatility evolves. We also use the more general RV-based
measure which fits an AR(1) process directly onto realized volatility.
Table 2: The Impact of Increases in Absolute Forecast Error on Confidence Intervals

<table>
<thead>
<tr>
<th></th>
<th>Δ CI Width</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Abs. Error Increased</td>
<td>1.65***</td>
<td>0.33</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>(2.84)</td>
<td>(0.56)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.77**</td>
<td>0.27</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>(-1.99)</td>
<td>(0.67)</td>
<td>(-0.49)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Control: Unexpected Volatility</td>
<td>None</td>
<td>GARCH</td>
<td>RV</td>
</tr>
<tr>
<td>Control: Expected Change in Volatility</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Observations</td>
<td>1,875</td>
<td>1,875</td>
<td>1,875</td>
</tr>
</tbody>
</table>

*t statistics in parentheses. *, **, *** 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 to calculate the absolute forecast error for this period. The eight-quarter-ago forecast is required to calculate the absolute forecast error from four quarters ago. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.
### Table 3: The Impact of Missing the Confidence Interval on Confidence Interval Widths

<table>
<thead>
<tr>
<th></th>
<th>( \Delta \text{ CI Width} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Miss CI</td>
<td>5.21***</td>
</tr>
<tr>
<td></td>
<td>(11.97)</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.33***</td>
</tr>
<tr>
<td></td>
<td>(-9.84)</td>
</tr>
<tr>
<td>Economic Significance</td>
<td>33.4%</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y</td>
</tr>
<tr>
<td>Unexp. Vol.</td>
<td>None</td>
</tr>
<tr>
<td>Exp. ΔVol.</td>
<td>None</td>
</tr>
<tr>
<td>Observations</td>
<td>4,643</td>
</tr>
</tbody>
</table>

\( t \) statistics in parentheses. *, **, *** 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. Economic Significance is the estimated change in CI width divided into the average initial CI width of 15.6%. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

### 5.1. Increases or Decreases in Absolute Forecast Error

Before testing the implications of our Bayesian learning model, we assess whether the data are consistent with the simple GARCH(1,1) model. Table 2 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 percentage points. Given that the average CI is 15.6% wide, this represents a modest 11% widening.

By controlling for unexpected volatility using either the GARCH or RV specifications, we no longer estimate a significant effect of the increase in absolute forecast error on changes in interval widths. This is consistent with periods of high unexpected volatility yielding wider intervals regardless of absolute forecast error. These results are inconsistent with the GARCH model presented above. Our evidence therefore suggests that forecasters are not updating their volatility forecasts using the rule from the simple GARCH model.
5.2. Hitting or Missing the CI and Distance from CI

Table 3 presents estimates from regressing changes in CI width on an indicator that activates when forecasters miss their confidence intervals. Column (1) presents estimates with no controls. In columns (2) and (3), we control for both unexpected volatility in the current period and the expected change in volatility using our GARCH- and RV-based model of volatility, respectively.

Without any controls, the estimated effect of missing the confidence interval is widening the subsequent forecast interval by a significant 5.2 percentage points. This is an economically significant amount as it represents widening the confidence interval by approximately 33%. Controlling for volatility, this remains large, between 4.8 – 5.2 percentage points.

The negative estimate of the intercept suggests that upon hitting the confidence interval, forecasters significantly narrow their confidence intervals. While our framework implies that the interval should narrow, we know on average that CFOs’ intervals are far too narrow. Essentially, hitting the interval appears to induce further overconfidence.

Overall, these results indicate that when forecasters miss the CI, the new interval widens by a large amount. We call this learning. However, as discussed in section 4, forecasters’ intervals are far too tight relative to the distribution of historical returns. Further, forecasters remain miscalibrated throughout our entire sample. 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. We return to this topic in section 6.

5.3. Asymmetries: High v. Low and Upper v. Lower

Table 4 shows results from regressing the total change in CI width on two indicators. The first activates when the realized return was higher than the forecaster’s upper bound. We call this missing the interval high. Similarly, if the realized return was lower than the lower bound of the confidence interval, we say that the forecaster missed the interval low, and the second indicator activates. If the forecaster hit the interval, both indicators are set to zero.

With no controls, forecasters who miss the interval high widen their confidence intervals by 25% or approximately 3.9 percentage points. Forecasters who miss low widen by 53% or approximately 8.3 percentage points. The estimates that control for volatility show a similar asymmetry between missing high or low. Forecasters who miss high widen by 27 – 30%, while those who miss low widen by 42 – 44%.

We also analyze how the upper and lower portions of the interval change when forecasters hit or miss their CIs. We define the upper portion of the CI (UCI) as the distance between
Table 4: The Impact of Missing the Interval High or Low on Confidence Interval Length

<table>
<thead>
<tr>
<th></th>
<th>Δ CI Width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Miss CI High</td>
<td>3.92***</td>
</tr>
<tr>
<td></td>
<td>(8.59)</td>
</tr>
<tr>
<td>Miss CI Low</td>
<td>8.25***</td>
</tr>
<tr>
<td></td>
<td>(14.81)</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.39***</td>
</tr>
<tr>
<td></td>
<td>(-10.12)</td>
</tr>
<tr>
<td>Economic Significance (Miss High)</td>
<td>25.1%</td>
</tr>
<tr>
<td>Economic Significance (Miss Low)</td>
<td>52.9%</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y</td>
</tr>
<tr>
<td>Unexp. Vol.</td>
<td>None</td>
</tr>
<tr>
<td>Exp. ΔVol.</td>
<td>None</td>
</tr>
<tr>
<td>Observations</td>
<td>4,643</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses. *, **, *** 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. Economic Significance is the estimated change in CI width divided into the average initial CI width of 15.6%. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.
Table 5: The Impact of Missing the Interval on Upper and Lower Confidence Intervals

<table>
<thead>
<tr>
<th></th>
<th>( \Delta \text{ UCI} )</th>
<th>( \Delta \text{ LCI} )</th>
<th>( \Delta \text{ UCI} )</th>
<th>( \Delta \text{ LCI} )</th>
<th>( \Delta \text{ UCI} )</th>
<th>( \Delta \text{ LCI} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss CI</td>
<td>(1) 2.17***</td>
<td>(2) 3.04***</td>
<td>(3) 2.20***</td>
<td>(4) 3.04***</td>
<td>(5) 1.97***</td>
<td>(6) 2.79***</td>
</tr>
<tr>
<td></td>
<td>(10.51)</td>
<td>(9.27)</td>
<td>(10.52)</td>
<td>(9.10)</td>
<td>(9.51)</td>
<td>(8.46)</td>
</tr>
<tr>
<td>Constant</td>
<td>(7) -1.45***</td>
<td>(8) -1.88***</td>
<td>(9) -1.31***</td>
<td>(10) -1.70***</td>
<td>(11) -1.27***</td>
<td>(12) -1.65***</td>
</tr>
<tr>
<td></td>
<td>(-9.01)</td>
<td>(-7.39)</td>
<td>(-8.22)</td>
<td>(-6.67)</td>
<td>(-7.91)</td>
<td>(-6.48)</td>
</tr>
<tr>
<td>Economic Significance</td>
<td>36.5%</td>
<td>31.5%</td>
<td>37.0%</td>
<td>31.5%</td>
<td>33.0%</td>
<td>28.9%</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Unexp. Vol.</td>
<td>None</td>
<td>None</td>
<td>GARCH</td>
<td>GARCH</td>
<td>RV</td>
<td>RV</td>
</tr>
<tr>
<td>Exp. ( \Delta Vol. )</td>
<td>None</td>
<td>None</td>
<td>GARCH</td>
<td>GARCH</td>
<td>RV</td>
<td>RV</td>
</tr>
<tr>
<td>Observations</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
</tr>
</tbody>
</table>

\( t \) statistics in parentheses. *, **, *** 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%. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

In column (1) of Table 5, we see that with no controls, forecasters who miss the CI widen the upper portion of their intervals by approximately 36% or 2.2 percentage points. The change in the upper portion of the interval accounts for approximately 42% of the total 5.2 percentage point widening of the CI upon missing the CI. From column (2), the remaining 58% of the total change is driven by widening the lower portion of the interval by 31% or approximately 3.0 percentage points. Controlling for volatility, we find similar results. Missing the CI widens the upper portion by approximately 33 – 37% or 2.0 – 2.2 percentage points, while the lower portion widens by 29 – 32%, or 2.8 – 3.0 percentage points.

In Table 6, we analyze changes in the upper and lower portions of the CI when forecasters miss high or low. Interestingly, our results indicate that forecasters who miss high widen the upper bound of the CI of the point forecast of the return. Similarly, the lower portion of the interval (LCI) is the distance between the point forecast and the lower bound of the CI.

We focus on how each portion of the CI changes but also present results on changes in the upper and lower bounds of the CI in Appendix A. We focus on changes in each part of the CI to ignore effects of the CI shifting solely because of the point forecast changing. We are interested in analyzing the shape of the CI and whether this changes, regardless of changes in the point forecast.

In column (1) of Table 5, we see that with no controls, forecasters who miss the CI widen the upper portion of their intervals by approximately 36% or 2.2 percentage points. The change in the upper portion of the interval accounts for approximately 42% of the total 5.2 percentage point widening of the CI upon missing the CI. From column (2), the remaining 58% of the total change is driven by widening the lower portion of the interval by 31% or approximately 3.0 percentage points. Controlling for volatility, we find similar results. Missing the CI widens the upper portion by approximately 33 – 37% or 2.0 – 2.2 percentage points, while the lower portion widens by 29 – 32%, or 2.8 – 3.0 percentage points.

In Table 6, we analyze changes in the upper and lower portions of the CI when forecasters miss high or low. Interestingly, our results indicate that forecasters who miss high widen the
Table 6: The Impact of Missing the Interval High or Low on Upper and Lower Confidence Intervals

<table>
<thead>
<tr>
<th></th>
<th>$\Delta$ UCI (1)</th>
<th>$\Delta$ LCI (2)</th>
<th>$\Delta$ UCI (3)</th>
<th>$\Delta$ LCI (4)</th>
<th>$\Delta$ UCI (5)</th>
<th>$\Delta$ LCI (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss CI High</td>
<td>1.76*** (8.09)</td>
<td>2.15*** (6.26)</td>
<td>2.19*** (9.86)</td>
<td>2.55*** (7.23)</td>
<td>1.96*** (8.95)</td>
<td>2.29*** (6.62)</td>
</tr>
<tr>
<td>Miss CI Low</td>
<td>3.13*** (11.77)</td>
<td>5.11*** (12.18)</td>
<td>2.25*** (7.16)</td>
<td>4.58*** (9.19)</td>
<td>2.02*** (6.18)</td>
<td>4.58*** (8.83)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.47*** (-9.17)</td>
<td>-1.92*** (-7.62)</td>
<td>-1.32*** (-8.18)</td>
<td>-1.84*** (-7.17)</td>
<td>-1.27*** (-7.85)</td>
<td>-1.82*** (-7.08)</td>
</tr>
<tr>
<td>Economic Significance (Miss High)</td>
<td>29.6%</td>
<td>22.3%</td>
<td>36.7%</td>
<td>26.5%</td>
<td>32.8%</td>
<td>23.8%</td>
</tr>
<tr>
<td>Economic Significance (Miss Low)</td>
<td>52.6%</td>
<td>53.0%</td>
<td>37.7%</td>
<td>47.5%</td>
<td>33.9%</td>
<td>47.5%</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y Y Y Y</td>
<td>Y Y Y Y</td>
<td>Y Y Y Y</td>
<td>Y Y Y Y</td>
<td>Y Y Y Y</td>
<td>Y Y Y Y</td>
</tr>
<tr>
<td>Unexp. Vol.</td>
<td>None None</td>
<td>GARCH GARCH</td>
<td>RV RV</td>
<td>RV RV</td>
<td>RV RV</td>
<td>RV RV</td>
</tr>
<tr>
<td>Exp. $\Delta$Vol.</td>
<td>None None</td>
<td>GARCH GARCH</td>
<td>RV RV</td>
<td>RV RV</td>
<td>RV RV</td>
<td>RV RV</td>
</tr>
<tr>
<td>Observations</td>
<td>4,643 4,643</td>
<td>4,643 4,643</td>
<td>4,643 4,643</td>
<td>4,643 4,643</td>
<td>4,643 4,643</td>
<td>4,643 4,643</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses. *, **, *** 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. 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%. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.
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, even accounting for a potentially higher point forecast, the forecaster puts a higher probability on a higher return realization. Similarly, upon missing low, the forecaster puts a higher probability on a lower return realization. Specifically, after controlling for volatility, forecasters who miss high widen their upper intervals by $30 - 37\%$ and their lower intervals by $22 - 26\%$. On the other hand, forecasters who miss low widen their lower intervals by $47 - 48\%$ and their upper intervals by $34 - 38\%$.

The evidence on asymmetry suggests that forecasters may be aware of whether the realized return was above, inside, or below their confidence intervals. 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 take this as further evidence of learning.

6. Learning and Miscalibration

In this section we analyze the cross section of individual forecasters. Our sample includes over 2,800 individual forecasters. We first drop all observations without a corresponding four-quarter-ahead observation from which we can observe learning. For the analysis that follows, we also drop all individuals with less than four matched-pairs of observations. This conservative precaution is to help ensure we are analyzing patterns across several repeated instances of observing respondents recalibrating. We are left with 361 forecasters who cumulatively made 3,180 forecasts.

For each of the observations in our sample of paired survey responses, we mark the respondent as having learned if (a) they missed the interval and subsequently widened their CI width or (b) they hit the interval and narrowed their CI width. We calculate the percentage of forecasters missing and learning as the number of times they missed and widened divided into the number of times they missed. Similarly, the proportion of hitting and learning is the number of times the forecaster hit and narrowed divided into the number of times they missed. We aggregate these two measures into a single measure of total learning.

Conditional on having at least four matched-pair observations, the median repeat forecaster has 7 responses, hits the interval in two of them, and misses in the other five. The median proportion of total learning is 50\% and this is balanced equally between learning from missing and learning from hitting. Figure 4 shows the distribution of learning among the 361 individual forecasters in our sample.\footnote{In work in progress, we are experimenting with a bootstrap simulation to determine whether what we call individual learning is different from random noise.}
Distribution of the proportion of learning among individual forecasters. A forecaster “learns” everytime they miss the interval and subsequently widen their CI or hit the interval and subsequently narrow their CI. The proportion of learning is the number of times a forecaster learns divided into their total number of forecasts.

We calculate each respondent’s individual calibration the same way we calculated mis-calibration per survey: dividing the number of responses that fell in the CI by the total number of responses\(^8\). The median forecaster’s calibration is 25.0%, well under the 80% we would expect. Of course, given that the median forecaster has only seven responses, it is not improbable that the calibration we calculate is below 80%. Restricting the sample to only those with more than 24 responses, the median forecaster’s calibration is 22.3%, which, surprisingly, is worse.

Table 7 presents estimates from regressing each forecaster’s total calibration on their proportion of learning. The estimates in this table show by how much a forecaster’s total

\(^8\)This measure of individual calibration includes only the repeat forecasts. The correlation between this measure of individual calibration and the measure constructed using all responses is 0.84.
Table 7: Regression of Calibration on Learning

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Learned</td>
<td>0.28***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Missed and Learned</td>
<td>0.32***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Hit and Learned</td>
<td></td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.42)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>14.66***</td>
<td>11.67***</td>
<td>31.50***</td>
</tr>
<tr>
<td></td>
<td>(4.23)</td>
<td>(3.80)</td>
<td>(12.85)</td>
</tr>
<tr>
<td>Observations</td>
<td>361</td>
<td>357</td>
<td>302</td>
</tr>
</tbody>
</table>

*t statistics in parentheses. *, **, *** denote significance at the 0.10, 0.05, and 0.01 levels under the assumption of a single test.

Regression of individual calibration on individual learning. Units are percentage points. The median individual calibration is 25%.

 "% Missed and Learned" is the number of times a forecaster missed the CI and widened their new interval divided into the number of times they missed. "% Hit and Learned" is similarly defined using the number of times a forecaster hit the CI and narrowed their new interval. "% Learned" aggregates learning across both misses and hits into one measure of overall learning.

Accuracy would improve if they were to change how often they learned. In columns (1) and (2) we see that forecasting improvement from learning primarily comes when CFOs miss and learn.

In particular, we see in column (2) that a forecaster’s calibration increases by 3.2 percentage points when their percentage of learning from missing increases by 10 percentage points. Given that the median forecaster misses five times and learns approximately half of the time, these estimates imply that one additional instance of missing and learning will improve total calibration by 6.4 percentage points.

6.1. Learning and Miscalibration

We have presented evidence that, as predicted by our models, forecasters learn from their mistakes when they miss the confidence interval. We have also shown that at the same time, even after repeatedly learning, forecasters remain extremely miscalibrated. We reconcile these two findings by documenting that although forecasters are learning, they do not learn by nearly enough to become calibrated.

For this analysis, we focus on the select subset of responses for which the forecaster missed the CI and subsequently widened their confidence interval. For comparison, we
construct a hypothetical counterfactual as follows. First, if the forecaster missed low, the new hypothetical lower bound is exactly the observed return, such that the return falls just inside the interval. Second, given this lower bound, we construct the upper bound such that the new distribution has the same measure of symmetry as the original distribution. Similarly, if the forecaster missed high, the hypothetical upper bound is exactly the observed return and the hypothetical lower bound is chosen to maintain the same asymmetry. This serves to scale the distribution while maintaining the original skew.

We then calculate the hypothetical change in the width of the confidence interval. Finally, we calculate by how much more or less the individual forecaster should have, on average, updated their interval relative to the counterfactual. Figure 5 plots the distribution of these differences across all forecasters. On average, respondents who missed the interval but learned would need to have widened their intervals by an additional 23.1 percentage points such that they would just hit the CI.

In Figure 6, we also calculate the average difference between actual and hypothetical width updating for all responses by survey. The periods for which forecasters generally performed worst are those for which the forecasters failed to update by the largest amounts. This implies that in those periods, not only were forecasters very poorly calibrated, but even upon learning, they failed to adjust by enough.

6.2. Individual Forecasters

In this section we analyze the behaviour of several forecasters for whom we observe a long series of forecasts. First, we highlight two forecasters whose behaviour is consistent with our models. We then show two forecasters whose behavior is less consistent with our models and discuss the causes and implications.

6.2.1. Forecaster #1363

Figure 7 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. For each survey date, the grey area represents the previously forecasted CI from one year ago. The red circle is the realized return and the black interval is the newly forecast CI for the one-year-ahead return.

Forecaster #1363’s behavior is largely consistent with the predictions of our Bayesian model. In the first four observations of Figure 7, the forecaster hits the CI and subsequently narrows his forecast. In the fifth and sixth observations, corresponding to 2007Q4 and

---

9The data begins in 2006Q1 because before that period, there were very few people in each survey for whom we could have observed learning at least four times.
The distribution across individuals for the mean difference between the hypothetical and actual changes in CI width. Units are percentage points. The hypothetical change is the change in the CI width that would have occurred had the forecaster constructed a hypothetical confidence interval (HCI) such that (a) the realized return just hit the HCI and (b) the HCI had the same symmetry as the actual confidence interval.

2008Q1, the forecaster hits the CI and widens their new interval. 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. 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.
The average difference between actual and hypothetical width updating for all responses by survey. The hypothetical change is the change in the CI width that would have occurred had the forecaster constructed a hypothetical confidence interval (HCI) such that (a) the realized return just hit the HCI and (b) the HCI had the same symmetry as the actual confidence interval.

6.2.2. Forecaster #2265

Figure 8 illustrates the evolution of Forecaster #2265’s 17 confidence intervals. We again find that the forecaster’s behavior is largely consistent with our model, 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, the forecaster becomes a “better” Bayesian learner. Early in the sample, the forecaster makes an incorrect inference (relative to our model) from the observed return falling outside of the CI, and widens the interval.
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.

instead of narrowing it. This is consistent with the forecaster incorrectly using Bayes’ rule to combine their prior belief with the observed return. Over time, and with experience, the forecaster becomes better at learning from observed returns, and hence their behavior is more consistent with our model.

6.2.3. 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 9, 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
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.

keeps the CI roughly the same or narrows it.

The forecaster then hits the CI three times from 2011Q3 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 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.

6.2.4. Forecaster #466

Forecaster #466 is an extreme example of the overconfidence of CFOs in our sample. The average width of this forecaster’s intervals is only 4.21%, just over one third of the average in the entire sample. While this may appear to be an outlier, approximately 23% of the intervals in our sample are 5% wide or less. Unsurprisingly, we see in figure 10 that this forecaster rarely ever hits the CI. The forecaster sometimes follows the prediction of our model and widens their CI in response to missing, but even then, the widening is by far too little to have a meaningful difference. This forecaster consistently produces very tight intervals, regardless of their poor forecasting performance.
6.3. Understanding the Persistence of CFO Miscalibration

Our paper presents a puzzle: even after repeated forecasts, CFOs are still badly miscalibrated. Although CIs increase after missing, the changes are a fraction of the width required for proper calibration. Our empirical analysis characterizes the miscalibration and allows for logical control variables but cannot answer the question of why.

Given that we have the identity of these CFOs, we are considering directly interviewing a subset of CFOs and asking them how they think about the forecast exercise. For example, we are very interested in talking to forecaster #466 who, as we showed in section 6.2.4, is both badly miscalibrated and very overconfident.

We have not yet formulated the interview questions. Following our past work that includes interviews, we know it is crucial to start with open-ended questions. These questions
are then followed by more specific questions. We are currently seeking input on these questions. Below, we include a first and very preliminary draft of the interview script.

[Introduction. Thanks for participating in the survey.]

**Q0.** Tell me about your background. [Here try to elicit educational background. Do they have an MBA, is their undergraduate quantitative?]

**Q1.** In the most recent survey, you forecast that the S&P500 would rise 10% over the next year and there was a one-in-ten chance the market would be higher than 20% and a one-in-ten chance lower than 0%. Can you talk to us about how you came up with the forecast and the range?

**Q2.** [Assuming most of the discussion is on the 10% forecast.] We are particularly interested in the range of the CI. In your latest survey, it is 20%. How did you come up with that?

**Q3.** [Potentially get another read of overconfidence]. Most consider the stock market to be related to the state of the economy. What is your forecast for real GDP growth over the next year? Please answer, in percent:

1. There is a one-in-ten chance that real GDP growth will be above ____.
2. Real GDP growth over the next year will be ____.
3. There is a one-in-ten chance that real GDP growth will be below ____.

[Survey instrument might have another test of overconfidence based on psychology literature.]

**Q4.** Are you aware of the most recent (approximate) value of the VIX (the instrument that represents the implied volatility on the S&P 500)? [Most should know this. If they don’t, do we tell them the value is 15% “which is historically consistent with the annual volatility of the S&P 500 stock returns”.]

**Q5.** If the stock return annual volatility is approximately 15%, how does this relate to your 80% range of 20%? [Here it will likely be revealed that there is some misunderstanding of the link between confidence intervals and standard deviations. There are many branches possible.]

**Q5a.** [Example Answer: 20% range makes sense because it is larger than the volatility which is 15%.] We usually think of a forecast and a confidence interval associated with that forecast. The usual 95% confidence interval is the forecast ± two standard deviations (± 30% given today’s volatility) and the 66% interval is ± one standard deviation (so ± 15% given today’s volatility). However, all of this is based on market data. Your interval is 20%, which is more precise than the market interval. We are interested in understanding the precision. Will you comment? [Do they believe they know more than the market or the average?]

**Q5b.** [Example Answer: The range I gave was for my forecasts not market outcomes.] The questionnaire says there is a one-in-ten chance the return will be greater or lower than ____. This refers to actual realized returns. What would your interval be on realized returns?
Q5c. [Example Answer: I know more than the average.] Volatility changes through time. [Try to extract information on the respondent’s model of how volatility changes and how their forecasts change.]

Q6. [Extract information on missing high vs. missing low in a hypothetical situation.]

We recognize that any CFOs interviewed as part of this analysis will need to be excluded (or, at the very least, treated differently) in future research on overconfidence.

7. Conclusion

[To be added.]
References


Kuchler, Theresa, and Basit Zafar, 2017, Personal experiences and expectations about aggregate outcomes, *FRB of NY Staff Reports*.


Murphy, Kevin, 2007, Conjugate bayesian analysis of the gaussian distribution., Technical report.


Table A1: The Impact of Missing the Interval on Upper and Lower Bounds

<table>
<thead>
<tr>
<th></th>
<th>$\Delta \text{UB}$</th>
<th>$\Delta \text{LB}$</th>
<th>$\Delta \text{UB}$</th>
<th>$\Delta \text{LB}$</th>
<th>$\Delta \text{UB}$</th>
<th>$\Delta \text{LB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss CI</td>
<td>(1) 4.57***</td>
<td>(2) -0.64*</td>
<td>(3) 4.37***</td>
<td>(4) -0.87**</td>
<td>(5) 4.30***</td>
<td>(6) -0.45</td>
</tr>
<tr>
<td></td>
<td>(16.14)</td>
<td>(-1.66)</td>
<td>(15.15)</td>
<td>(-2.21)</td>
<td>(15.14)</td>
<td>(-1.16)</td>
</tr>
<tr>
<td>Constant</td>
<td>(1) -3.23***</td>
<td>(2) 0.10</td>
<td>(3) -3.02***</td>
<td>(4) 0.00</td>
<td>(5) -3.00***</td>
<td>(6) -0.08</td>
</tr>
<tr>
<td></td>
<td>(-14.67)</td>
<td>(0.33)</td>
<td>(-13.70)</td>
<td>(-0.01)</td>
<td>(-13.63)</td>
<td>(-0.27)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Unexp. Vol.</td>
<td>None</td>
<td>None</td>
<td>GARCH</td>
<td>GARCH</td>
<td>RV</td>
<td>RV</td>
</tr>
<tr>
<td>Exp. $\Delta$</td>
<td>None</td>
<td>None</td>
<td>GARCH</td>
<td>GARCH</td>
<td>RV</td>
<td>RV</td>
</tr>
<tr>
<td>Vol.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses. *, **, *** 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 (LB) bounds of the confidence interval on an indicator that activates when the realized return missed the confidence interval. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year-ago realized volatility.

Appendix A. Further Results

We also analyze how the upper and lower bounds of the interval change when forecasters hit or miss their CIs. Table A1 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 3.

For example, recall that without controls, missing the interval widened the CI by approximately 5.2 percentage points. The estimate in column (1) shows that most of this change is driven by an increase in the upper bound. Similarly, when controlling for volatility, the majority of the total change is driven by increasing the upper bound.

Table A2 presents estimates from regressing the changes in the upper and lower bounds on indicators that activate when forecasters miss the interval high or low. Forecasters who miss high raise their upper bounds by 5.0 – 5.7 percentage points. At the same time, however, forecasters raise their lower bounds by 0.9 – 1.3 percentage points, partially offsetting some of the widening effect from increasing the upper bound. The increasing of the CI is consistent with shifting the entire interval upward in response to a return realization above the forecasted CI.

Forecasters who miss low primarily adjust by lowering their lower bounds. With no controls, we estimate that missing low decreases the lower bound by almost 4.8 percentage points. Controlling for volatility, missing low decreases the lower bound by 6.5 – 6.9
Table A2: The Impact of Missing the Interval High or Low on Upper and Lower Bounds

<table>
<thead>
<tr>
<th></th>
<th>∆ UB (1)</th>
<th>∆ LB (2)</th>
<th>∆ UB (3)</th>
<th>∆ LB (4)</th>
<th>∆ UB (5)</th>
<th>∆ LB (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss CI High</td>
<td>5.03***</td>
<td>1.11***</td>
<td>5.65***</td>
<td>0.90**</td>
<td>5.57***</td>
<td>1.32***</td>
</tr>
<tr>
<td></td>
<td>(16.81)</td>
<td>(2.79)</td>
<td>(18.91)</td>
<td>(2.23)</td>
<td>(19.09)</td>
<td>(3.31)</td>
</tr>
<tr>
<td>Miss CI Low</td>
<td>3.50***</td>
<td>-4.75***</td>
<td>0.33</td>
<td>-6.50***</td>
<td>-0.30</td>
<td>-6.90***</td>
</tr>
<tr>
<td></td>
<td>(9.57)</td>
<td>(-9.78)</td>
<td>(0.78)</td>
<td>(-11.32)</td>
<td>(-0.68)</td>
<td>(-11.55)</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.21***</td>
<td>0.18</td>
<td>-2.66***</td>
<td>0.50*</td>
<td>-2.57***</td>
<td>0.53*</td>
</tr>
<tr>
<td></td>
<td>(-14.61)</td>
<td>(0.62)</td>
<td>(-12.27)</td>
<td>(1.68)</td>
<td>(-11.84)</td>
<td>(1.78)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Unexp. Vol.</td>
<td>None</td>
<td>None</td>
<td>GARCH</td>
<td>GARCH</td>
<td>RV</td>
<td>RV</td>
</tr>
<tr>
<td>Exp. ΔVol.</td>
<td>None</td>
<td>None</td>
<td>GARCH</td>
<td>GARCH</td>
<td>RV</td>
<td>RV</td>
</tr>
<tr>
<td>Observations</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
<td>4,643</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses. *, **, *** 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 (LB) bounds of the confidence interval on indicators that activate when the forecaster missed the interval high or low. All regressions include fixed forecaster effects. Unexpected volatility is the difference between realized volatility and either the GARCH- or RV-based forecast of volatility. Expected change in volatility is the difference between forecast of volatility and one-year ago realized volatility.

percentage points and yields no significant change in the upper bound.

Appendix B. Proofs

Let the $t = 0$ denote the prior and $t = 1$ denote the posterior time periods. The prior belief $\gamma_0 = \beta_0/\alpha_0$ while the posterior belief $\gamma_1 = (\beta_1 \epsilon_{t-1}^{-2})/\alpha_1$. Thus:

$$\gamma_1 > \gamma_0$$

$$\frac{\beta_1 \epsilon_{t-1}^{-2}}{\alpha_1} > \frac{\beta_0}{\alpha_0}$$

$$\frac{[\beta_0 \epsilon_{t-1}^2 + \frac{1}{2}(r_t - \bar{r})^2] \epsilon_{t-1}^{-2}}{\alpha_0 + \frac{1}{2}} > \frac{\beta_0}{\alpha_0}$$

$$\beta_0 + \frac{1}{2}(r_t - \bar{r})^2 \epsilon_{t-1}^{-2} > \beta_0 + \frac{1}{2} \frac{\beta_0}{\alpha_0}$$

$$(r_t - \bar{r})^2 \epsilon_{t-1}^{-2} > \frac{\beta_0}{\alpha_0}$$

$$(r_t - \bar{r})^2 > \frac{\beta_0}{\alpha_0} \epsilon_{t-1}^{-2}$$

$$(r_t - \bar{r})^2 > \sigma_t^2$$