NBER WORKING PAPER SERIES

RATIONALIZING FIRM FORECASTS

Nicholas Bloom
Mihai A. Codreanu
Robert A. Fletcher

Working Paper 33384
http://www.nber.org/papers/w33384

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

We would like to thank the Kauffman Foundation and Stanford University for their extensive financial support. We would like to thank seminar participants at the AEA, AFE, Atlanta Fed, Chinese University of Hong Kong, Census Bureau, Chicago, Chicago Harris, Clemson, Columbia, Davis, EMC, ERMAS, EEA-ESEM, EFA, HBS, Maryland, Microsoft, NBER Entrepreneurship, NBER Organizational Economics, Northwestern, Kellogg, Santa Clara, Stanford, UCSC, USC, UVA, WEAI, and Wharton. We would also like to thank Mert Akan, Elias Bruegmann, Patrick Collison, Eilin Francis, Kevin Didi, Yue Duan, An He, Nate Hilger, Ana-Maria Mocanu, Alex Novet, Derek Ozkal, Bo Xu, Ethan Yeh, and David Yuan for their feedback and support, and to Bill Aulet, Camille Herbert and David Thesmar as our conference discussants. This paper was preregistered as AEARCT R-0007058 (https://www.socialscienceregistry.org/trials/7058). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Rationalizing Firm Forecasts
Nicholas Bloom, Mihai A. Codreanu, and Robert A. Fletcher
NBER Working Paper No. 33384
January 2025
JEL No. J0

ABSTRACT

We conduct a five-year, ten-wave panel survey in collaboration with a large U.S. payments company, collecting incentivized sales forecasts from over 6,000 firms. The sample is representative of the size distribution of U.S. employer firms. We find that firms exhibit remarkably poor forecasting accuracy, with bias, predictable errors, and over-precision, particularly in smaller firms. To explore the sources of forecast inaccuracy we implement four randomized controlled trials: improving data use, increasing incentives, providing training, and encouraging contingent thinking. While enhanced data use and incentives lead to modest improvements in contemporaneous forecast accuracy, training and contingent thinking yield no measurable effects. None of the interventions improve forecasting accuracy in subsequent waves. Surveying economists, we find they predicted about half the inaccuracy and bias we found, highlighting the surprisingly large deviation of firms from rational expectations.

Nicholas Bloom
Stanford University
Department of Economics
and NBER
nbloom@stanford.edu

Mihai A. Codreanu
Stanford University
Department of Economics
mihaic@stanford.edu

Robert A. Fletcher
Stanford University
robertsfletch@gmail.com

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/7058

I. Introduction

Firms' expectations play a central role in economics, from Keynes's General Theory to models of investment under uncertainty, business cycles or search and matching. These models usually assume firms have rational forecasts. A fast-growing behavioral literature, however, highlights systematic departures from rationality including over-optimism, slow updating and over-precision. Despite this we still have scant empirical evidence on the accuracy of firms' forecasting. There are two key challenges in the existing literature: (i) sample representativeness and (ii) data quality.

First, existing studies often focus on large listed firms, leaving the majority of businesses unexamined. There are 7 million employee firms in the US, but only 4,000 of these (less than 0.1%) are publicly listed (U.S. Census Bureau, 2023). As Figure 1 shows US listed firms are far larger, with a median of almost 1,500 employees compared to 3 for the average U.S. employer firm. These size differences matter: large firms often rely on professional forecasting teams, whereas smaller firms typically rely on owner-generated forecasts. For example, a survey of 1,139 US firms by the Atlanta Federal Reserve Bank in 2025 found that 45% of firms with 500+ employees had professional sales forecasters while only 7% of those with 1-9 employees did (Appendix Figure A1). And while large firms are important for aggregate output, firms with less than 100 employees are central to economic dynamism, accounting for 87 percent of establishment entry and exit, and 46 percent of net job creation (U.S. Census Bureau, 2022). A representative analysis of firm expectations should thus include the full-size distribution of firms.

Second, it is also hard to obtain and evaluate sales forecast data from firms. Most surveys provide little or no incentive for truthful, considered responses, inviting concerns over inattention, careless errors or strategic reporting (e.g., Bertrand and Mullainathan, 2001). Public forecasts—such as earnings guidance—may be carefully prepared, but can be strategically biased to influence investors. Moreover, it is often difficult to evaluate sales forecasts against future realizations. Publicly listed US firms provide annual sales data, but as noted these comprise less than 0.1% of all US employee firms. Private firms rarely provide accounting data, so evaluating their forecasts usually requires re-surveying them in subsequent years. But this resurvey data can itself be biased or noisy (e.g. Altig et al. (2020)) making it hard to evaluate if sales forecast errors come from poor forecasts or noisy follow-up surveys.

To address these challenges, we undertook two steps: First, we partnered with one of the largest U.S. financial services company ("FinTech"), which serves over 1 million U.S. businesses. Using FinTech's transaction-level data, we analyzed 6,594 firms, stratified to the U.S. employee firm-size

distribution. Second, we conducted a 10-wave, five-year panel survey, offering each respondent a $50 base payment, a $25 per-round participation fee, and up to $400 for accurate forecasts (defined as predicting next-quarter revenue within ±10 percent of actuals). Collecting and evaluating these forecasts was expensive, with over $1.1 million in payments given out to firms. But it yielded the first large, incentivized, US panel of firm forecasts.

Our first key result is that firms' forecasts are highly inaccurate. Only 18 percent were within ±10 percent of actual realizations. We identify three key problems driving this low accuracy.

First, bias. On average, forecasts are biased, with a mean over-optimism bias of 16 percent. More importantly, this bias has a wide idiosyncratic variation. In a sample of firms responding 8 or more times to our survey, firm fixed effects explain more than 30 percent of forecast errors. Second, predictable errors. Forecast errors are correlated with lagged errors, past growth, and past sales. Third, over-precision. The survey elicited not only point-forecasts but also second-moment forecasts using 3 and 5-bin outcomes and probabilities. Firms provided subjective forecast variances that were about a half of the true variance, highlighting they were over-confident in their forecasting accuracy.

One immediate question is why are firms forecasts so inaccurate? Perhaps our reward payments were not enough to elicit high efforts from firms. Or perhaps business owners are profit maximizing, with over-optimism a way to motivate employees and investors? Alternatively, biases could be costly, and maybe they arise from poor management practices or the inability of individual owners to make good forecasts?

To investigate these questions, we ran four randomized controlled trials. First, we randomized firms' exposure to a data-dashboard on their recent FinTech revenue to evaluate the extent to which reviewing historical sales data improves forecasting. Second, we randomized accuracy rewards, ranging from $0 to $400, to evaluate to what extent greater incentives for accuracy improved forecasting. Third, we randomized forecast training, providing up to 10 vignettes of a forecasting exercise for a hypothetical business to see if forecasting improved with experience. Finally, we randomized firms into evaluating scenarios in advance of making their forecasts, building on the "Superforecaster" literature. This literature suggests good forecasters consider a wide range of possible outcomes, including best and worst case scenarios in advance of making forecasts (e.g. Tetlock and Gardner, 2015).

We found these randomized controlled trials had a limited and temporary impact on forecasting accuracy. The data dashboard exposure reduced bias and over-precision, and increased win rates by about 2 percentage points. However, its effects did not persist into future waves. This was surprising as we estimate the returns to dashboard use just from survey rewards - so ignoring any

potential additional business benefits - was about $500 per hour. One explanation for this bias is firms are too confident: they believe they know more than they do and do not think external aids can improve their performance, a common trait seen in studies of individual forecasters (e.g. Kahneman, 2011).

The reward randomization reduced average optimism bias. Firms randomized into the highest $400 reward buckets saw mean optimism bias drop by around three quarters. But since the major source of inaccuracy was idiosyncratic bias and forecasting error, this caused only a limited improvement in forecasting accuracy. Win rates only increased by 0.9 percentage points for every additional $100 in rewards. This improvement also did not persist into future rounds. This suggests some optimism bias is pliable – firms are aware of it, and if adequately incentivized can reduce their bias – but this effect is only transitory. One explanation is motivated reasoning, in that owners are deliberately over-optimistic to motivate themselves, employees or investors (e.g. Bénabou and Tirole, 2002).

Finally, neither the forecast training nor the contingent thinking interventions had any impact on bias, error predictability, over-precision or the win rates. This suggests firm forecasting inaccuracy does not arise from a lack of forecast training or consideration of tail outcomes.

In the last empirical part of the paper, we examine the prior beliefs that economists have on both the forecasting accuracy of firms as well as the effectiveness of our experiments. Economists estimated inaccuracy, bias and over-precision of only about half the level seen in our sample. And just like the authors of this paper, they were also over-confident on the impact of experiments, expecting improvements in win rates from the interventions that were two or more times higher than we observed in the data. One explanation is that the bulk of previous literature focuses on forecasting in sophisticated large firms. Putting these biases into an illustrative quantitative model we find these forecast biases generate over-investment, excess volatility and increased misallocation.

The paper relates to five literatures.

First, there is a literature evaluating the biases of firm managers. Malmendier and Tate (2005, 2008) use data on public stock holdings of Forbes 500 CEOs to show they hold excessive levels, implying over-optimism. More recent papers directly measure firm sales forecast data in large firms (e.g., Graham, Harvey and Puri, 2013; Gennaioli, Ma and Shleifer, 2016; Barrero, 2022; Ma et al., 2024). These studies find moderate overconfidence and predictable errors in large firm samples with median sales ranging from $10M to $1B – equivalent to the top 1 percent to 5 percent of U.S. firms.[1] These biases are part of the recent push away from traditional rational expectations models (Gabaix, 2023; Moll, 2025). We show that forecasting errors are substantially bigger in our

[1]Values in 2022 US dollars to match the Census 2022 Business Dynamics Statistics (U.S. Census Bureau, 2022).

firms. One explanation is large firms typically have professional forecasters while in our firms, like in most US firms, forecasts are usually made by the owner.[2]

Second, there is a literature on management practices, showing how firms have widely varying practices, with smaller firms typically having worse management practices (e.g. Bloom et al., 2019). This literature finds poor management practices are hard to persistently change and have a large negative impact on firm performance (Bruhn, Karlan and Schoar, 2018). Some of the issues with poor management arise from problems analyzing data including forecasting (e.g. Gosnell, List and Metcalfe, 2020; Cai and Wang, 2022). These practices appear worse in developing countries where firms are smaller (e.g. McKenzie, 2018; McKenzie and Sansone, 2019). Our paper ties in closely with this literature, highlighting poor forecasting is likely a worse issue in developing countries, where firms are usually smaller than US firms.

Third, our results echo the findings from a long-standing literature in psychology on biases. One strand shows individuals often have optimistically biased forecasts (e.g. Weinstein, 1980; Moore and Healy, 2008). Another strand documents over-precision in that individuals believe their forecasts are more precise than they are (e.g. Fischhoff, Slovic and Lichtenstein, 1977; Soll and Klayman, 2004). Finally, a third strand highlights how individuals fail to learn from mistakes, so can persistently be over-optimistic despite incurring costs from this bias (e.g. Sharot, Korn and Dolan, 2011; Kahneman, 2003). These biases are all present in our firms. One explanation is in our firms forecasts are made by the individual owner, so these forecasts inherit the biases usually found in studies on individuals forecasting.

Fourth, the behavioral economics literature has attempted to rationalize these biases. One explanation of over-optimism is as a type of motivated belief which managers find useful for motivating themselves, employees or investors (e.g. Bénabou and Tirole, 2002; Landier and Thesmar, 2009; Bénabou and Tirole, 2016). One famous example is Steve Jobs' "Reality Distortion Field" in which he maintained excessively over-optimistic beliefs to motivate employees and investors (Isaacson, 2011). The need to maintain motivated beliefs over time may lead agents to become over-confident, so they fail to learn from mistakes (e.g. Zimmermann, 2020; Huffman, Raymond and Shvets, 2022).

Finally, there is a literature exploring firm and professional forecasters' expectations about macroeconomic variables like GDP or inflation. This is a long literature and includes papers by Coibion and Gorodnichenko (2015), Bordalo et al. (2020) and Farmer, Nakamura and Steinsson (2024). In these models, forecasts are rationally made, but with errors arising from the constraints

[2]Another literature, including Ben-David, Graham and Harvey (2013), Tanaka et al. (2019), Boutros et al. (2024) and Coibion (2018); Coibion, Gorodnichenko and Ropele (2020) collects data on firms forecasts on external variables like the S&P500, GDP or inflation. Bachmann and Elstner (2015) collect qualitative firm sales forecasts.

of sticky information (e.g. Mankiw and Reis, 2002) or noisy information (Sims, 2003).

More broadly our research also contributes to the literature using randomized controlled trials for assessing and mitigating biases (e.g. List, 2011; Floyd and List, 2016). This literature is limited in using managers as an experimental group in part due to difficulties of accessing such samples and in part due to the difficulties (or high costs) in eliciting effort from them. We partnered with FinTech to obtain size representative samples of US firms and used over $1.1 million in financial rewards to elicit effort.

Our paper continues as follows. Section 2 explains our survey design and data collection, Section 3 discusses our forecast biases, while Section 4 discusses the results of our randomized controlled trials. Section 5 examines the prior predictions by economists of firms' forecasting biases as well as an illustrative model of the importance of these biases, while Section 6 concludes.

II. Data

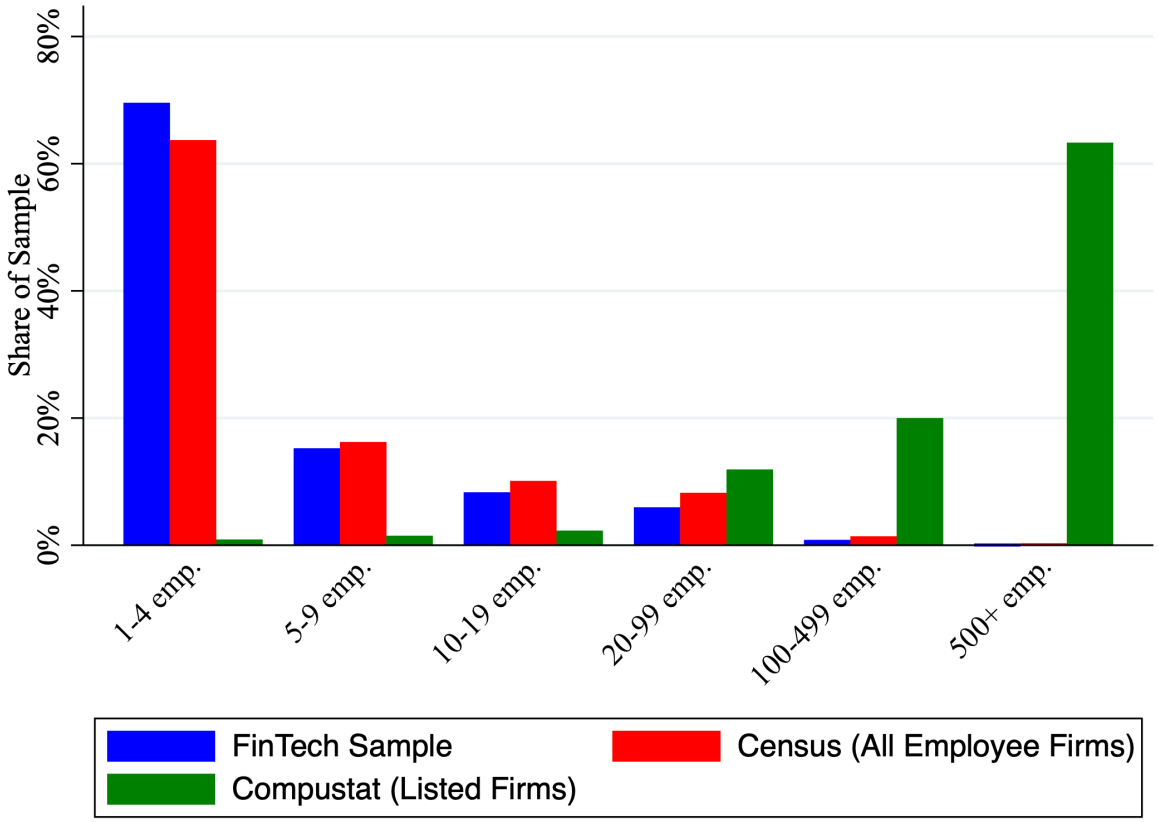
A. Survey Design and Details

We partnered with a large international payments technology firm operating in the United States, henceforth referred to as FinTech. They operate both online and in person payments systems (some examples are shown in Appendix Figure A3). The baseline sample of firms was drawn from their universe of US businesses with active accounts. Businesses had to be for-profits, and the contact emails had to be individuals listed as account owners.[3] To be eligible to enter the sample, firms had to have set-up an electronic payment processing option (e.g., accept card payments), which is very common in the US (for example, the Federal Reserve Financial Services (2024) estimates this is 86 percent of firms). We sent 10 waves of surveys over 5 years from 2019 to 2024, with the sample for wave 1 in spring 2019 consisting of about 18,000 firms, stratified to match the size distribution of US firms. Surveys were sent out about every 6 months, to allow 3 months between each survey waves to assess firms' 3-month survey forecasts and pay any rewards before the next survey invitation. From wave 2 onwards we invited all firms that responded in the prior waves, and a booster set of firms that did not complete the survey before, under similar random stratified sampling conditions. The boosters were added in waves 3, 5, 7, and 9, inviting approximately 35,000 additional firms. Due to data access restrictions, we can only report results and collect the data on the firms that answered our survey.[4]

The surveys were targeted at business owners, obtaining a sample of approximately 94 percent of responses coming from an original founder or current CEO. Throughout the survey waves, response

[3]For example, their emails could not consist of generic phrases such as info@, admin@, or contact@.
[4]This is due to data use consent restrictions.

Figure 1. : Employment Size Comparison: Our Sample, Census and Compustat



Note: This figure shows the share of firms by the number of employees in the FinTech, Census and Compustat samples. The FinTech sample contains the 6,594 firms with 18,060 observations in our survey. The Census data is extracted from the Country Business Register for 2021, that is available at https://www.census.gov/data/datasets/2021/econ/susb/2021-susb.html. The Compustat compiles responses from 2019-2024, containing 26,830 industrial firm responses that were active in the US, had non-missing non-zero employees, sales, plants and equipment net of depreciation.

rates hovered at about 10–15 percent. Due to data access restrictions, we cannot present the results of the likelihood to respond based on survey invitations. What we can do is present robustness for all results using inverse probability weights by attrition rates throughout the panel.[5]

Each survey wave consisted of a core set of questions on the finances and current performance of the business, labor use, and forecasting. On top of this, there is a rotating series of modules asking about one-off topics such as management, COVID-19, personality traits, macroeconomic forecasting, capital and other shorter points of interest. Finally, waves from five onwards included experimental modules, all detailed in Table 1.

[5]Attrition was positively related to revenue and negatively related to employees (Appendix Table A1), but overall with a small explanatory power (R-squared).

Firms were contacted with an invitation e-mail and three follow-ups spaced five days apart. The compensation was formed from two distinct parts: a flat participation fee and an accuracy reward. The flat fee was $50 for responding to the first wave of the survey and then $25 for each subsequent wave. The accuracy reward was a $25 per survey wave bonus if their next quarter revenue forecast was within 10 percent of their realized revenue. This $25 accuracy reward was increased to up to $400 as part of our rewards randomization interventions in waves 5, 7 and 9.

To ensure participation, we kept the surveys short such that they can be completed in 10–12 minutes. Nonetheless, we received a sizable number of responses from multi-million-dollar companies (winsorized at the 97.5th percentile). Across all 10 waves of the survey we received 18,060 valid responses. Throughout our data collection effort, we paid out around $1.1 million in forecast payments and accuracy rewards. More details of the data sampling, structure, cleaning, as well as more detailed experimental design can be found in the Data Appendix B.

Table 1—: Project Timeline

Wave	Dates	Responses	Extra Module	Randomized Control Trial
1	Jan-Apr 2019	2785	Baseline	
2	Jun-Aug 2019	1685	Management	
3	Oct 2019 - Jan 2020	1994	Personality	
4	Apr-May 2020	1602	COVID-19	
5	Sep-Oct 2020	1765	COVID-19	Dashboard, Reward
6	Jan-Apr 2021	1254	Management	Dashboard
7	Sep-Oct 2021	2760	Forecasting	Dashboard, Reward, Training
8	Mar-Jun 2022	1272	Forecasting	Dashboard, Scenarios
9	Oct 2022-Jan 2023	1735	Forecasting	Dashboard, Reward, Training
10	Nov 2023-Jan 2024	1208	Funding	Dashboard, Training, Scenarios

Notes: This table reports the timeline of the 10 waves of the Project (column 1), by the main outer window dates the wave responses/predictions starts across all waves and rounds (column 2), and total number of clean survey responses (column 3), totaling 18060. The cleaning procedure from all raw responses to the final dataset is described in the Data Appendix. Each wave had an extra module (column 4), and starting wave 5 at least one randomized control trial for the incentivized 3-month sales forecasting questions (column 5).

B. Business Characteristics

The characteristics of the firms and CEOs are reported in Table 2. The firms have a mean revenue of $752,000 a year, with 6 employees. They are relatively online focused with 70 percent of their revenue online and just below 60 percent of total revenue via FinTech.

The firms span the entire United States with coverage across almost all states and industries (Appendix Figure A4). The average firm completed 3 survey waves, with 106 firms completing all

10 waves. The CEOs are typically 39 years old, with 41 percent female. They are well educated, with 95 percent having some college and 31 percent some graduate education.

C. Payment Data

The partnership with FinTech has two big advantages. First, we are able to sample a set of firms much more representative to the US firm size distribution than previous studies focusing on Compustat-size firms. Second, using the FinTech data we can precisely match firms answering our survey with their aggregate revenue, total transactions and average transaction value. This enables us to accurately assess high frequency firm responses against actual revenue data. This information is also readily available to firms, at any time, via their FinTech account dashboard.

Using FinTech data also comes with a cost. In our case, we explicitly limited the predictions to FinTech revenue so that we could directly verify their forecasts. We found that because of the different transaction fees, timing and customer types, firms track revenues by channel.[6] To further ensure our results are not driven by heterogeneity in FinTech usage, we also ask firms about their total revenues and share of FinTech revenues and, throughout the paper, examine this heterogeneity.

III. Firm Sales Forecasts

In each wave of the survey, respondents were asked to predict their FinTech revenue for the next three months and the next twelve months.[7] The surveys were designed to be short: in the piloting process, the first-round duration was about 15 minutes if taken in one sitting and the follow-up surveys could be completed in about 10 minutes without interruptions.

In addition to participation fees, respondents were promised an Amazon gift card reward[8] if their three-month prediction was within 10 percent of their actual revenue. Prior to the fifth wave, this reward was a fixed value of $25 to ensure respondents had a financial incentive to provide considered responses. In waves 5, 7, and 9, we randomized the reward amounts offered to respondents (see Appendix Figure A10). Specifically, 25 percent of respondents were randomly assigned to receive no reward, another 25 percent continued with the base amount of $25, and the remaining 50 percent were randomly placed into eight equally sized groups (6.25 percent each) corresponding

[6] We conducted about 20 (qualitative) interviews with firms in our sample in the early waves of the study. We found that firms care about their channel of revenues. For example, cash payments have no payments fees while some online vendors charge up to 30 percent. Similarly, vendors have very different chargeback and fraud management policies, as well as settlement frequency (i.e. time the money gets from the customer to the business account).

[7] While we only formally incentivized the 3-months predictions, the forecast errors between the 3-month and subsequent 12-month forecasting exercises had a linear correlation of .68 (t-stat of 119.4).

[8] In later rounds, we used a provider that allowed easy redemption of the gift card at many other retailers, including Amazon.

Table 2—: Firm Descriptive Statistics

	Observations	Mean	Median	Std. Dev.
Panel A: Firms				
Revenue last 12 months ($'000s)	6594	752	72	2581
Number of Employees	6316	5.6	2	8.5
Revenue Online, %	6594	70.3	90	35.7
Revenue from FinTech, %	6594	59.5	61	31.2
Number of Survey Rounds	6594	2.74	2	2.3
Panel B: CEOs				
Business Founder	6594	0.94	1	0.24
Age (years)	6121	39.4	38	10.5
Female	6145	0.41	0	0.49
1 to 3 Years College	6141	0.19	0	0.39
4 Years College	6141	0.45	0	0.50
Masters +	6141	0.31	0	0.46
STEM College Degree	6054	0.33	0	0.47
Panel C: Treatments				
Dashboard RCT	9994	0.49	0	0.50
Reward RCT	6260	0.74	1	0.44
Training RCT	5703	0.34	0	0.48
Scenarios RCT	2480	0.47	0	0.50
Panel D: Forecasts				
Winner	18060	0.180	0	0.38
Forecast Error (arc-percent)	18060	0.163	0	0.89
Squared Forecast Error	18060	0.812	0	1.24
Over-Precision	9810	0.394	0	0.49
Variance (Subjective Distribution)	9810	0.275	0	0.56

Notes: This table reports descriptive statistics for 6,594 unique firms (totaling 18,060 responses). The table includes information at entry for the first two subpanels and for all valid responses for the last two. The variables are: all firm earnings in the last 12 months, number of employees (including full and part-time), percentage of revenue online, percentage of revenue on FinTech, and the number of survey waves with valid responses; CEO is a business founder, age of the CEO, self-reported gender of the CEO is female, highest level of education attained by the CEO, and an indicator if a college degree in a STEM field was obtained; dummy variable indicators of being treated by the dashboard experiment (before prediction), reward experiment (reward different than $25), training and scenario experiments (before prediction) in the relevant waves; being classified as a winner, forecast error (bias) calculated as 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months), squared forecast error, over-precision is the share of the firms that had an actual revenue either above their highest scenario or below their lowest, and variance calculated using the algorithm in Appendix A1.3. All employee, earnings, and age variables were winsorized at the 2.5th and 97.5th percentiles of the distribution in the first survey wave they responded.

to reward values of $50, $100, $150, $200, $250, $300, $350, and $400. The reward amounts were independently randomized in each wave and independent of other treatments.

We also asked firms to provide their subjective earnings distribution. We elicited these via 3-bin (lowest, central, highest) or 5-bin (lowest, low, central, high and highest) sales forecasts and their corresponding probabilities such that they add to 100 percent (see question in Appendix Figure A12, data details in Appendix B). This follows the approach in Altig et al. (2020) to collect data on business level sales uncertainty, yielding a subjective probability distribution for future sales.

Forecast error is measured using the arc-percent difference between the predicted revenue and the actual revenue. The denominator is the average of the predicted and actual revenue[9] to constrain the range to [-2,2] to limit the impact of outliers, and to avoid large values for near-zeros in predicted or actual revenue:

$$\text{Forecast Error}_{it} = \frac{\text{Predicted}_{it} - \text{Actual}_{it}}{0.5 \times \text{Predicted}_{it} + 0.5 \times \text{Actual}_{it}} \tag{1}$$

A. Forecasting Importance

Before analyzing firms' forecasting accuracy, it is worth confirming that forecasting is important for firm performance. There is a long history in economics arguing that if firms face adjustment costs —such as for investment, hiring, pricing, or marketing—they need to forecast future demand.[10] This seems to be confirmed by firm self-reports. In wave 8 we also polled firms on the importance of forecasting for deciding on hours, number of employees, materials, advertising, and capital expenditures in the business. On average, 76.4 percent of firms said forecasting is one of the top three most important factors for one or more of these decisions (Appendix Figure A5). Some other (free text box) responses included forecasting as being important for pricing, product releases or branch closure. Hence, businesses perceive forecasting accuracy as important for performance.

To quantify importance empirically, we see that firms with greater forecast accuracy are more successful in being larger, faster growing, more likely to survive and have a higher management score (Appendix Figure A6).[11] Forecasting errors also predict employment, with firms that are over-optimistic in prior rounds having significantly more employees both in the current and in future waves (Appendix Table A4). This is suggestive of labor misallocation. We see that even after controlling for time, industry and even firm fixed effects, forecast errors are predictive of a

[9]This measure is common in the literature and was adapted from the Davis and Haltiwanger (1992) growth rate measure.

[10]Some classics in this literature include Hayashi (1982), Bertola and Caballero (1994), Dixit and Pindyck (1994), Abel and Eberly (1996) and Cooper and Haltiwanger (2006). All involve firms maximizing profits by forecasting future demand or productivity in the presence of adjustment costs. This is separate from another reason to forecast sales accurately which is related to borrowing constraints or fundraising.

[11]Putting this differently, Figure A6 also shows descriptive evidence of the negative relationship between firm size and accuracy, as we asserted in the introduction.

bigger future firm size. Of course, these estimates focus on labor, the most direct channel we can measure. However, as shown in Appendix Figure A5, forecasting is also perceived as important when deciding materials, advertising, and capital expenditures (with materials expenditures most sensitive to forecasting).

B. Baseline Forecasting Structure

Before discussing results we provide a framework to characterize the firm's forecasting behavior. This allows us to decompose forecast errors into average and idiosyncratic bias, and define predictable errors and over-precision.

Let A_{it} be the actual revenue of firm i in quarter t. The true, underlying process for revenue can be written as:

$$A_{it} = f(\mathcal{I}_{t-1}) + \epsilon_{it} \tag{2}$$

where $f(\mathcal{I}_{t-1})$ represents the systematic, predictable component of revenue based on all information available at time $t-1$, $\mathcal{I}_{t-1}$. The term ϵ_{it} is the unpredictable shock to revenue, with $E[\epsilon_{it}|\mathcal{I}_{t-1}] = 0$. Under rational expectations, a firm's forecast, $P_{it}^{RE} = E[A_{it}|\mathcal{I}_{t-1}] = f(\mathcal{I}_{t-1})$ However, firms in our sample may deviate from this rational benchmark. We model the firm's actual point forecast, P_{it}, as incorporating a systematic error, Ψ_{it}, which represents all non-rational components of the forecasting process. Defining the total forecast error u_{it}=$\Psi_{it} - \epsilon_{it}$ (a more general term of the empirical arc-percent error in Equation 1) we can write:

$$P_{it} = P_{it}^{RE} + \Psi_{it} = f(\mathcal{I}_{t-1}) + \Psi_{it} = A_{it} + u_{it} \tag{3}$$

This decomposition clarifies that the observed forecast error u_{it} is driven by two unobservable components: the firm's predictable error, Ψ_{it}, and the mean zero unpredictable shock, $-\epsilon_{it}$. The forecasting issues we document are as follows:.

Bias. This occurs when the $E[\Psi_{it}] \neq 0$. This contains three distinct components:

$$\Psi_{it} = \underbrace{B}_{\text{Avg. Bias}} + \underbrace{b_i}_{\text{Idiosyncratic Bias}} + \nu_{it} \tag{4}$$

Here, B is the average bias across all firms, b_i is a firm-specific persistent bias (some firms can be systematically optimistic, others pessimistic). ν_{it} is a residual shock to the prediction.[12]

[12] As we see in the data the systematic error can have a time varying θ_t, due to unforseen macro-economic shocks, in particular the COVID pandemic and rebound in our data, even if $E_{t-1}[\theta_t] = 0$.

Predictable Errors. Forecast errors can be predictable using information available at the time the forecast is made. For example, in a regression of the form:

$$u_{it} = \lambda u_{it-1} + \gamma' X_{it-1} + v_{it} \tag{5}$$

where X_{it-1} includes variables like past growth or sales, then $\lambda \neq 0$ or $\gamma \neq 0$ indicates that errors are predictable. One reason for predictability is from idiosyncratic bias in Ψ_{it} in the error term.

Over-precision. This occurs when firms believe their forecasts are more accurate than they are. We assess this by comparing a firm's subjective uncertainty to the objective uncertainty of its outcomes. Let $\text{Var}_s(u_{it})$ be firms subjective expectation of the variance of the probability distribution over future sales, elicited from our 3- and 5-bin questions. Let $\text{Var}(u_{it})$ be the true variance of the firm's forecast errors. Over-precision is the condition where a firm's subjective uncertainty is systematically lower than the observed forecast error variance:

$$\text{Var}_s(u_{it}) < \text{Var}(u_{it}) \tag{6}$$

This implies that firms' subjective distributions are too narrow, leading to frequent "surprises" where realized outcomes fall outside the range of what firms considered possible.

C. Forecasting Issues

We quantify the three key forecasting issues, using the framework of the forecasting structure in Equations 2-8. We discuss each of them in turn.

A) Bias:

Mean Bias B. The average forecast error is +16.3 percent, representing a large positive mean bias. The presence of over-optimism bias has been shown before in the literature (Manski, 2018; Altig et al., 2020; Barrero, 2022; etc.). However, these results come from uncompensated surveys or public forecasts where other incentives may bias predictions.[13] In a sample of firms with more than half of their income on FinTech - a sample which averages 82% of their revenue coming from FinTech - the average over-optimism slightly increases to +16.4 percent. In a sample of firms with more than 99% of their revenue on FinTech it is +15.3 percent. Table 3 shows that over-optimism is a persistent feature of our firms across the business cycle. Our survey spanned from 2019—a period of stable expansion—through the deep COVID-19

[13]While we technically elicit something close to the mode of the distribution, in practice (as can be seen in later sections, e.g. the spread of the subjective distributions in the 3-bin and 5-bin forecasts) for most firms, the mode, mean and median forecasts are not meaningfully different. Under normality assumptions, these fully coincide.

recession of 2020, the recovery of 2021 and 2022, marked by significant macroeconomic uncertainty, and the post-pandemic stabilization of 2023 and 2024. Across all 10 waves of the survey the responses were significantly over-optimistic, as shown in column (2).

Idiosyncratic b_i. There is also a large substantial idiosyncratic bias. Figure 2 plots the mean bias for firms that made 8 or more forecasts. This is the counterpart of the distribution of terms b_i from Equation (6). Because of the panel structure of our data we are able to measure this. The mean bias ranges from -37 percent at the 10th percentile to 51 percent at the 90th percentile, highlighting how both over-pessimistic and over-optimistic firms exist in this sample. Within this sample this idiosyncratic bias accounts for 30.1 percent of the variance in the forecast error.[14] In the sample of firms that is also conditioned to have more than half of their income on FinTech, the idiosyncratic bias accounts for 34.9 percent of the variance in the forecast error.[15] In Appendix Figure A8 we have different iterations of the same figure, using different persistent respondents responses (for 5+, 6+, 7+ and 9+ responses), and we show that the R-squared for these samples ranges between 29.1 percent and 32.4 percent.

Time Varying Component of $\nu_{i,t}$. Forecast errors contain both idiosyncratic and time varying components. We note that contractionary periods had over-optimism rates between 4–11 percent, while expansionary periods had higher over-optimism rates, ranging between 22–35 percent. These are not biases, and ex-ante their expectation (net of mean bias and idiosyncratic bias) should be zero. Moreover, the magnitude of these in explaining forecast errors is not large. In a regression of the 18,060 observation sample, the R-squared of the time-varying fixed effects is only 2.1 percent. In a specification including firm fixed effects the introduction of time fixed effects increases the R-squared by just 0.9 percentage points. In the sample of firms with more than half of their income on FinTech, time fixed effects account for 2.8 percent of the total variance, while the R-squared increases from a regression of just firm fixed effects by 1.4 percentage points. The relevant numbers for the even more restricted sample of firms having more than 99% of their revenues on FinTech are 5.2 percent and just below 3 percentage points.

B) Predictable Errors: Firms' forecast errors are also predictable. As Figure 3 and Table 4 show, current forecast errors are highly correlated with lagged forecast errors[16] and other known

[14] Calculated as the R^2 on the firm fixed-effects in a regression of forecast errors for firms with 8+ responses (p-value <0.001).

[15] Conditioning on the share of firms with more than 99% on FinTech was 47.8 percent, although this was on a small sample size since it conditions on both more than 8 responses as well.

[16] The positive relationship between current and future forecast errors implies an under-reaction of firms to previous errors. We note that both the slope and the R^2 of this relationship is similar quantitatively to Ma et al. (2024); Thesmar (2025).

Table 3—: Mean Forecast Errors Over Time

Wave	Dates	Obs	Unweighted Mean	Reweighted Mean
(1)	(2)	(3)	(4)	(5)
1	Jan-Apr 2019	2785	0.35***	0.35***
2	Jun-Aug 2019	1685	0.22***	0.25***
3	Oct 2019-Jan 2020	1994	0.29***	0.31***
4	Apr-May 2020	1602	0.06***	0.09***
5	Sep-Oct 2020	1765	0.08***	0.11***
6	Jan-Apr 2021	1254	0.08***	0.10***
7	Sep-Oct 2021	2760	0.05***	0.06***
8	Mar-Jun 2022	1272	0.11***	0.13***
9	Oct 2022-Jan 2023	1735	0.04**	0.05**
10	Nov 2023-Jan 2024	1208	0.27***	0.32***
All	Jan 2019-Jan 2024	18060	0.16***	0.17***

Notes: This table reports statistics on mean forecast errors, calculated as 2*(Forecast-Actual)/(Forecast+Actual). Columns 4 and 5 show the unweighted mean and the weighted mean reweighted to account for survey attrition. The inverse probability weights are derived from the regression model in Appendix Table A1, column (4), using lagged sales, employment and industry fixed effects. T-tests for the null hypothesis that the mean is zero are performed and significance is denoted. *, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.

variables like sales growth. Because of the need to evaluate forecast accuracy in wave t and pay firms their reward before starting wave t+1, there is a gap of about 3 months between each survey wave, making this autocorrelation of forecast errors even more striking.[17] As we show in the last column of Table 4, these result are robust and statistically similar if introducing time, industry and also control for the share of revenues on FinTech.

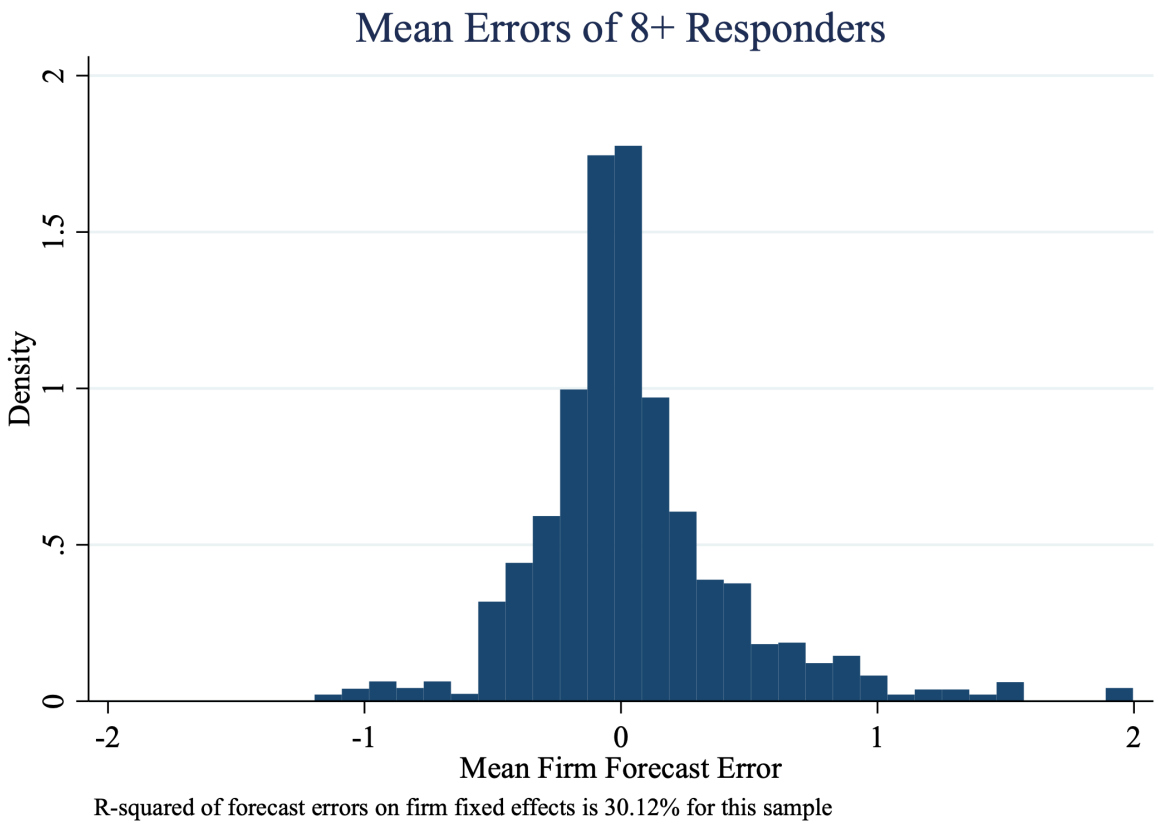
One explanation highlighted in Coibion and Gorodnichenko (2015) and Coibion (2018) is slow information updating due to rational inattention or information collection costs. This is likely one factor, but it is worth noting that as Table 4 shows firms' forecast errors are correlated with up to 5 lags of prior forecast errors stretching back more than 2 years.[18]

C) Over-precision: We also evaluate the higher moments of forecasts by eliciting distributions. Between waves 5–10, we asked firms for their 3-bin forecasts (lowest, middle and highest) or 5-bin forecasts (lowest, low, middle, high, and highest) for revenue in the next three months. We convert these into a subjective distribution of sales around their mean forecast using arc-percentile

[17]The exact timing between when surveys were run was also influenced by logistical constraints related to the FinTech's collaboration and permissions, as well as gift card acquisition and processing.

[18]Longer lags have positive (but declining) coefficients, but samples sizes are too small for lags beyond 5 to be significant. The declining lags are perhaps explained by the increased attention on the most recent information (Afrouzi et al., 2023).

Figure 2. : Mean Forecast Errors Have a Wide Variation Among Firms

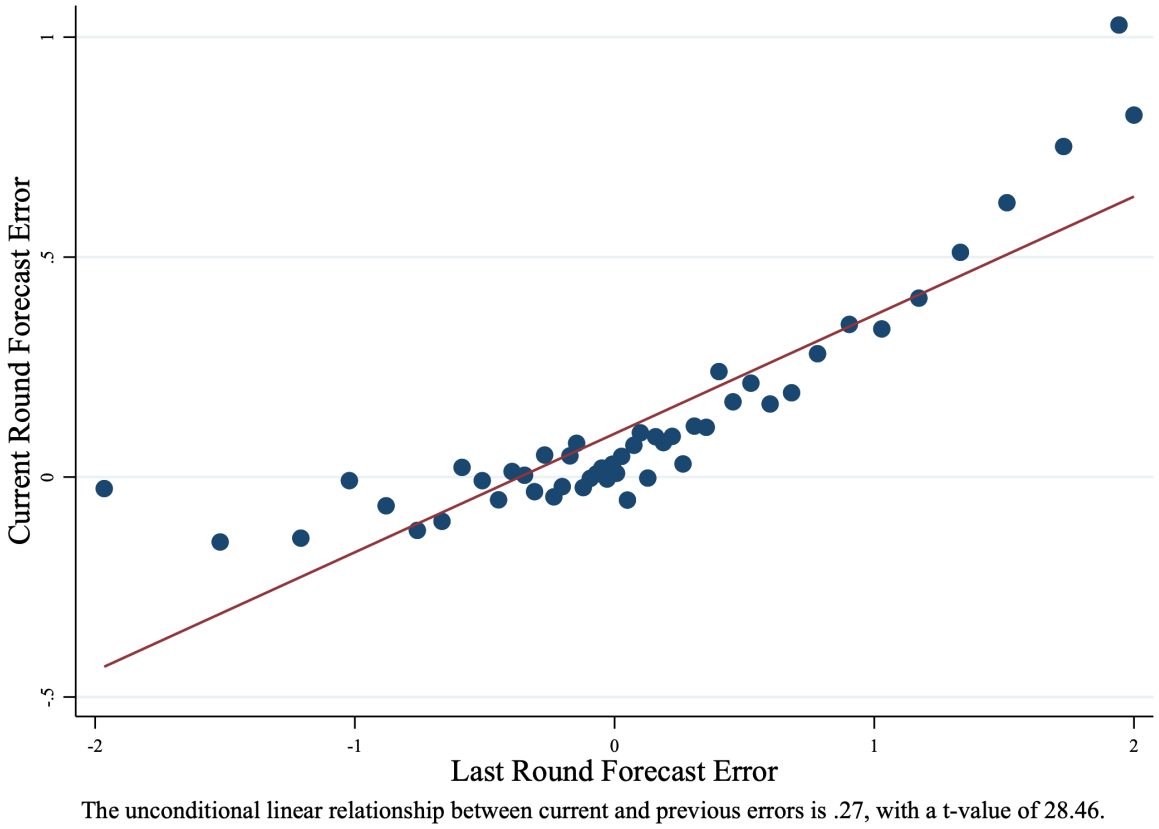


Note: This plot reports the average forecast error, for all firms that responded 8 or more times during our survey. Plot comprising 4,019 responses over waves 1-10, and 457 firms. The mean firm level bias is calculated by averaging the forecasting error (arc-percentage) across all forecasts. This is calculated using 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months). All firms have equal weight. For different iterations of the same figure, using different persistent respondents responses (for 5+, 6+, 7+ and 9+ responses), we show the mean firm forecast errors histograms in Appendix Figure A8.

variations (outcome-mean forecast)/(0.5*outcome+0.5*mean forecast). Figure 4 plots this as the red "subjective" distribution. We also plot in blue the distribution of actual sales realized in the next 3 months, again as an arc-percentile variation around the mean forecast.[19] Comparing these two distributions we see that the subjective distribution standard deviation is about 50 percent lower than the actual distribution, and its variance is just below half. Another measure of over-precision is the share of realizations that are outside of the (lowest, highest) interval. We find that

[19]For example, if a firm had a 3-bin forecast of next quarter's sales of 90, 100 and 110 with probabilities 0.25, 0.5 and 0.25 its mean would be 100. This would generate forecast variations around the mean of -10.5%, 0 and 9.5% calculated as $(90-100)/(0.5\times 90+0.5\times 100)$, $(100-100)/(0.5\times 100+0.5\times 100)$, and $(110-100)/(0.5\times 110+0.5\times 100)$. If the actual outcome was then 115 its value for the realization would be 13.9% calculated as $(115-100)/(0.5\times 115+0.5\times 100)$. More details for this procedure can be found in Appendix B.B3.

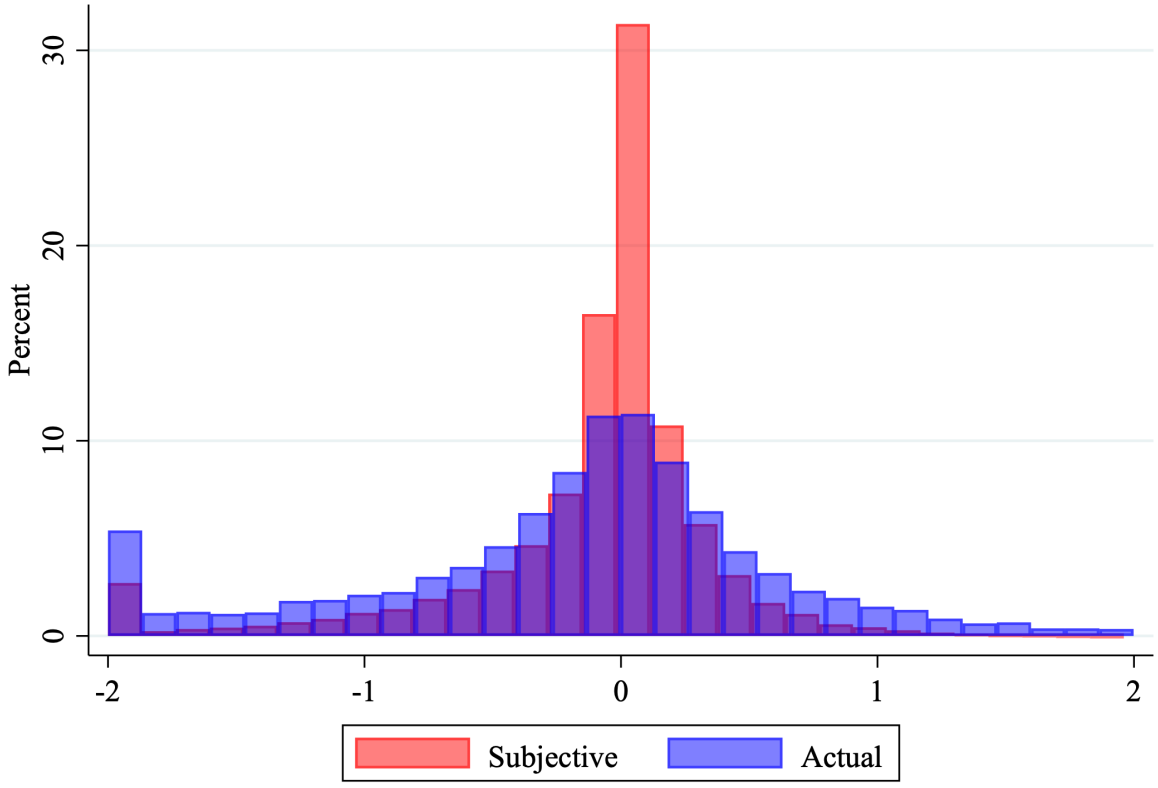
Figure 3. : Forecast Errors are Predictable



Note: Plot comprising 9,116 responses over waves 2-10, of 3,052 firms that have at least one valid previous wave response. The forecasting error (arc-percentage) is calculated using 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months), for the firm sales revenues 3 months forward from the relevant month either of the current wave or the prior wave with a valid firm response. Red line represent a linear line of best fit.

39.4 percent of realizations are above the firms' highest forecasts or below their lowest forecasts. For the sample of firms with more than half of their revenue on FinTech, the percent of realizations outside of these bounds is 38.8 percent while for firms with more than 99% of their revenues on FinTech it is 37.4 percent. So these results are also robust to conditioning on more exclusive FinTech usage.

Figure 4. : Subjective forecast distributions are over-precise (too narrow)



Note: Plot showing the subjective and actual distributions of sales, normalized by mean forecasts. Both calculated as an arc-percentage deviation, defined for actual as (Actual Sales Next 3 Months –Mean Forecast Next 3 Months)/(0.5*Actual Sales Next 3 Months+0.5*Mean Forecast Next 3 Months) and for subjective for each bin that receives a weight according to the subjective self-assessed probability distribution, as (Forecast Bin Sales Next 3 Months - Mean Forecast Next 3 Months)/(0.5*Forecast Bin Sales Next 3 Months +0.5* Mean Forecast Next 3 Months). Subjective taken from the 3- and 5- bins of forecasted sales and weighted according to forecast probabilities. Data from periods 5-10 for 9,810 responses and 4,251 firms. More details in the Data Appendix.

Table 4—: Predictable Errors

	(1) Forecast Error	(2) Forecast Error	(3) Forecast Error	(4) Forecast Error	(5) Forecast Error	(6) Forecast Error	(7) Forecast Error	(8) Forecast Error
Forecast Error$_{i,t-1}$	0.260***	0.206***	0.200***	0.238***			0.240***	0.234***
	(0.018)	(0.019)	(0.022)	(0.026)			(0.026)	(0.026)
Forecast Error$_{i,t-2}$		0.198***	0.173***	0.134***			0.135***	0.132***
		(0.017)	(0.020)	(0.023)			(0.023)	(0.023)
Forecast Error$_{i,t-3}$			0.082***	0.060**			0.059**	0.058**
			(0.020)	(0.024)			(0.024)	(0.024)
Forecast Error$_{i,t-4}$				0.087***			0.085***	0.082***
				(0.022)			(0.022)	(0.022)
Last Q Growth$_{i,t-1}$					-0.047***		-0.072***	-0.066**
					(0.017)		(0.027)	(0.027)
Log(1+Last Q Sales)$_{i,t-1}$						-0.045***	-0.005	-0.011
						(0.005)	(0.007)	(0.007)
Wave by Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE & Fintech Share	No	No	No	No	No	No	No	Yes
F-test (equal lags p-value)		0.782	0.000	0.000			0.000	0.000
R-squared	0.075	0.106	0.109	0.131	0.017	0.031	0.137	0.143
Observations	9108	6522	4490	3018	9108	9108	3018	3018

Notes: This table reports the relationship between forecast errors and pre-determined variables. The dependent variable is the forecast error, defined as 2*(Forecast - Actual)/(Forecast + Actual). Columns (1)-(4) regress forecast errors on their own lags. Column (5) uses lagged growth, Column (6) uses lagged size. Column (7) and (8) use all of the previous independent variables, with Column (8) also controlling for industry fixed effects and the share of the revenue coming from FinTech. The F-test reports the p-value for the null hypothesis that all lagged forecast error coefficients are equal. Stars denote significance: *** p<0.01, ** p<0.05, * p<0.1.

Overall, these forecasting issues led to low accuracy. In our sample of 18,060 predictions only 18 percent of firms receive their forecasting reward for being within 10 percent of their next quarter's sales. This forecasting accuracy is lower than even simple forecasting metrics like using each firms' last quarter sales.[20] Given firms have a large amount of additional information above their last quarter sales – such as imminent orders, new products, seasonality and customer flow – they should be able to do substantially better.

D. Forecasting Accuracy and Size

One important contribution of our paper is that we sample a set of firms that are more representative of the full US firm-size distribution, which has the median firm in the 1-4 employees bin. In Table A3, we formally test whether bigger firms indeed forecast differently than smaller firms. We regress bias (the forecast error), accuracy (mean squared error) and win rates on the log of employment. We find that size is indeed strongly predictive of accuracy and win rates in our sample, even when introducing industry or FinTech revenue share controls. Surprinsigly, size is not predictive of average bias. If anything, the point estimates suggest that bigger firms in our sample may be more over-optimistic. This may suggest that over-optimism may not be as detrimental to firm survival and performance. One potential explanation is that firm owners may work harder or pour more resources into their business when over-optimistic.

IV. Experiments

There are several potential explanations for these forecasting issues. One is informational, in that perhaps firms are making optimal forecasts but with stale or incomplete information. Another is that firms may be optimizing a different objective function from forecast accuracy. For example, perhaps for motivation reasons they are deliberately over-optimistic and the rewards in our experiment may not be high enough to justify paying attention costs. Otherwise competition could drive out firms that are not making optimal forecasts.[21] A third explanation is perhaps some firms lack forecasting skills, so cannot understand our survey questions or are unable to compute forecasts. A fourth explanation could be that firms may neglect to consider the complete range of possible scenarios, overly focusing on the most likely outcome.

To evaluate these explanations in turn, we ran four experiments on our firms, randomizing data

[20]Note this average win rate of 18% is increased by two of our experiments that did improve forecast accuracy (data and rewards). The untreated firms had a win rate of 17.4% in this period, in the absence of interventions. In the same period, the untreated firms would have had about 1.9 percentage points higher win rates if their forecast was the last quarters' sales instead.

[21]These firms can survive indefinitely with poor forecasts because owners can subsidize them with additional labor and capital, which aligns with the literature showing low returns to entrepreneurship (e.g. Moskowitz and Vissing-Jorgensen (2002))

access, financial incentives, training and scenario evaluations. We discuss these four experiments one by one, ordering them in the sequence we ran them chronologically.[22]

A. Information: Randomizing Usage of the Data Dashboard

As part of using FinTech to process their payments, businesses are given access to a financial dashboard. This dashboard (shown in Appendix Figure A9) appears when users open their account, and is free and easily accessible at the top of their account deck at any point. The dashboard allows users to see various financial metrics, including revenue by various periods including by day, week, month and quarter.

One motivation for the information treatment was that while the average forecast error was 16.3 percent, in a simple decomposition the mean bias term accounts for less than 5 percent of the mean squared forecast error across the waves (mean squared forecast error across all waves was 0.812).[23] This aligns with the work of Kahneman, Sibony and Sunstein (2021) which shows the importance of noise in prediction errors in various domains, arguing this issue is under-studied and suggesting more data could improve forecasting accuracy.

The intervention randomized if respondents were requested to open their dashboard immediately prior to being asked for their forecasts. This was done in each of waves five to ten, assigned with 50 percent probability in each wave (before invitation emails were sent) independently by wave and independently of all other interventions. Dashboard treated respondents were asked to log into their FinTech account and then enter: (A) the last 6-digits of their account number from their dashboard, and (B) choose the last three months in the dashboard dropdown menu and report their revenue from their dashboard.[24] The control group was only asked to report their revenue from the last three months with no suggestion to use their dashboard.

In Table 5 we report on the results of the dashboard experiment. In the top-panel (A) we start with survey diligence metrics – measures of how seriously respondents appear to take the survey question. Column (1) reports the time taken to complete the full survey. The dashboard intervention increases time spent by about 44.7 seconds, on a non-treated dependent variable mean of 1184 seconds. This indicates that the dashboard viewing and completing the required questions took approximately three-quarters of a minute. In column (2) we report the impact on the time taken on the prediction question, which increases by 7.6 seconds on a mean of 68 seconds, suggesting

[22]The first two experiments - the dashboard and rewards experiments - were preregistered (details in Appendix C) with the verbatim text. The other two experiments were added in later waves, so were not preregistered, although these were also evaluated on the same preregistered metrics and neither had any significant impact.

[23]Note that Mean Squared Error (MSE) is $\text{MSE} = \text{Bias}^2 + \text{Noise}$. Hence, we have the decomposition $\text{MSE} = \text{Bias}^2 + \text{Noise}$. At baseline the mean bias^2 is 0.027, which is 3.3 percent of the mean squared error of 0.812.

[24]We asked for the last 6-digits of their account as an inducement to use the dashboard since this is something managers typically would not know. We did not enforce completing correct 6-digits to continue answering the survey.

having additional data moderately increases how long they take to make their revenue forecasts. In column (3), we report the mean squared report error on the last three months' reported revenue, which shows a large 12.9 points reduction. This shows, perhaps unsurprisingly, that viewing the dashboard improves the accuracy of reporting their prior quarter revenue.[25] Finally, column (4) reports the share of digits that are rounded in their forecast as an indicator of careful forecast consideration. This is calculated as the share of "0" trailing digits in their forecast.[26] We see the dashboard lowered rounding by 2.2 percentage points.

Panel B examines forecasting outcomes. In column (1) we examine the mean forecast-error (bias) and find this drops by 3.7 points, suggesting more information also reduces over-optimism (decreasing it by about a third from the control group baseline). In column (2) we examine the squared forecast-error (noise) and find this drops by 9.7 points (decreasing it by just below 15% of the control group baseline).

Unsurprisingly given the results on bias and noise, in column (3) panel B we see the win rate rises by 2 percentage points on a base of 17.2 percent in the control group. Hence, survey participants randomized into the dashboard intervention won about 12 percent more frequently. Evaluating this in terms of return on time we know from Panel A the dashboard intervention increased the survey time by approximately 45 seconds. A conservative estimate would be that perhaps 20 seconds of this additional 45 seconds was for finding the last quarter revenue, while the rest was for completing the survey questions, and finding the firm ID. Under these assumptions and with a mean prize value in the dashboard waves of \$83.1, the estimated post-tax return of checking the dashboard for previous revenue is \$300 per hour.[27] This likely represents a conservative estimate of the perceived return, as most respondents vastly over-estimate their chance of winning (detailed below). Moreover, this is likely not taxed so is equivalent to perhaps \$500 an hour of pre-tax revenue.

In column (4) we examine the impact of the dashboard on error predictability. To do this we regress the forecast error on its lag, on the dashboard RCT and an interaction of these two variables, with the interaction reported. This shows the dashboard led firms to reduce their current forecast error coefficient on their lagged forecast error by about 0.05. While this is insignificant, the coefficient magnitude is about 18.5 percent of the mean value on lagged error of 0.27, providing some weak evidence that better information also reduces error predictability.[28]

[25]Note it does not reduce this to zero because we do not have 100 percent compliance with the dashboard intervention. And even for managers that looked at their dashboard they could still approximate, misremember or typo when entering the revenue.

[26]So, for example, a revenue forecast of \$62,500 would be 40 percent (two out of six digits are trailing "0"s), while \$650,000 would be 66 percent (four out of six digits are trailing "0"s), and \$603,105 would be 0 percent (none of the trailing digits are "0").

[27]This is calculated as the marginal increase in win rates, times the win reward amount, expressed in per hour earnings or, more formally: $0.02 \times 83.1 \times 3600/20$.

[28]Note that this is a significantly smaller sample for the error predictability regressions. We can only examine firms subject to the RCT in this column who have also made forecasts in a prior wave.

In column (5) of panel B we examine the impact of the dashboard experiment on the extent of over-precision, measured as the share of the forecasts outside the lowest and highest revenue forecasts from the 3-bin and 5-bin predictions. We see this fell by 3.3 percentage points, which is just below a tenth decrease given a base of 41 percent. This highlights the information intervention not only increased firms point-estimate precision but also improved higher-moments of their subjective forecast distributions, even though these estimates were not directly incentivized.

Figure 5 upper-panel shows the impact on the forecast-error for quantile regressions from the 5th to the 95th quantile. This reveals that the dashboard intervention raised the forecast for the lower percentiles of the forecast error (that were negative) and decreased it for the higher percentiles of forecast error (that were positive). Hence, the dashboard information improved accuracy by reducing the tail forecast errors.[29] Figure 5 bottom-panel shows the distribution of forecast errors, highlighting that the dashboard intervention reduced the dispersion, but the magnitude of the reduction is not large.

Surprisingly, given this improved forecast accuracy there is no dynamic effect. In panel C we regress each of our forecasting outcomes from panel B on the lagged dashboard intervention. This is a (smaller) sample of firms who respond in two consecutive surveys, so we can examine the impact of the dashboard randomization on their forecast in the subsequent period. All impacts are insignificant in this round and all future rounds.[30]

Collectively, the results of the dashboard intervention support models of inadequate data use by firms, perhaps because of firms' excessive confidence in their forecasting ability. Firms can substantially improve their win rates, earning perhaps $500 an hour from dashboard use, but fail to do so. This excessive confidence in their own forecasting ability aligns with the results from section III whereby managers over-estimate the precision of their forecasts. It also matches evidence from a question we asked respondents in wave 8 about their self-assessed probability of winning, in which they report a mean estimated win-rate of about 63 percent, more than three times their actual win rate.[31]

[29] Interestingly, we cannot reject that the treatment effect in the squared historic error and in the squared forecast errors is the same in absolute value. This may be connected to Huffman, Raymond and Shvets (2022) models of motivated beliefs in which individuals are motivated to distort memories of feedback to preserve unrealistic (e.g. over-optimistic) expectations. This is also related to the recall and bias effect in Taubinsky et al. (2025).

[30] This is of course subject to no impact of the randomization to future survey response rates by firm type. We formally test this in Appendix Table A2, and find that the dashboard does have a small impact on future response rates, but not any other experiments.

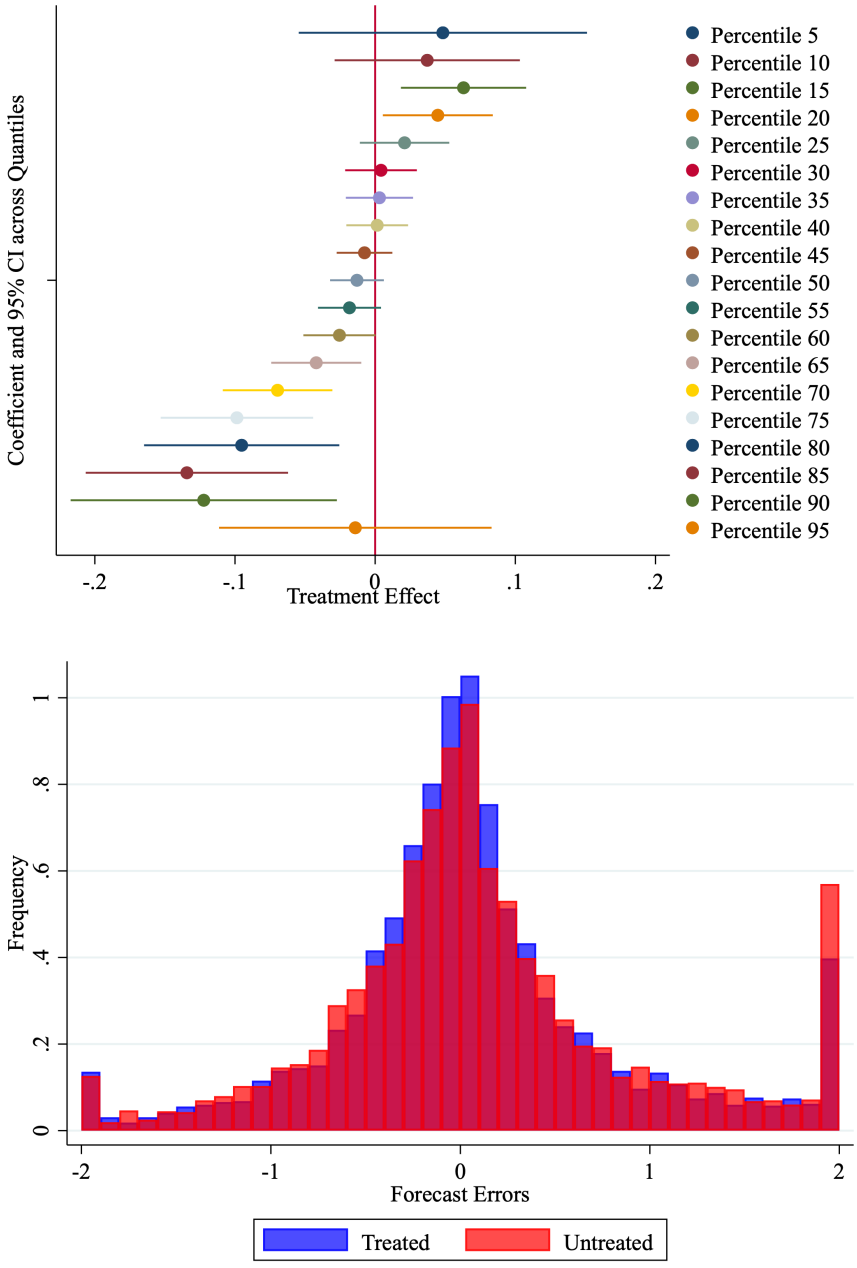
[31] The precise question is "How likely do you think you are to win, meaning your prediction is within 10% of your actual FinTech revenue in 3 months? Please response with a percentage between 0% and 100%".

Table 5—: The Dashboard Treatment

Panel A: Forecast Diligence					
	(1) Survey Time (seconds)	(2) Prediction Time (seconds)	(3) Reporting Error^2	(4) Rounding (% digits)	(5)
Dashboard$_{i,t}$	44.718***	7.557***	-0.129***	-2.203***	
	(9.632)	(2.455)	(0.019)	(0.449)	
N	9994	5837	9844	9994	
Dependent Mean	1184	68.0	0.453	58.3	
Panel B: Outcomes					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Dashboard$_{i,t}$	-0.037**	-0.097***	0.020**	-0.050	-0.033***
	(0.016)	(0.022)	(0.008)	(0.038)	(0.010)
N	9994	9994	9994	5594	9810
Dep. Mean	0.110	0.676	0.172	0.269	0.410
Panel C: Persistence					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Dashboard$_{i,t-1}$	-0.025	-0.034	-0.009	0.028	0.011
	(0.022)	(0.029)	(0.012)	(0.050)	(0.015)
N	4585	4585	4585	3153	4523
Dependent Mean	0.081	0.561	0.212	0.225	0.392

Notes: This table reports results of the dashboard treatment on survey and forecasting diligence (in Panel A), main forecasting outcomes (in Panel B), and persistence (in Panel C). The independent variable is a dummy taking the value 1 if a firm is treated by the dashboard, and 0 otherwise either this wave (Panels A and B) or last wave (Panel C). In Panel A, the dependent variables are: survey time, the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); prediction time, the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 10 minutes (column 2); historic sales squared error, calculated as [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 3); and rounding percentage of digits, the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panels B and C: forecast error (bias) which is calculated as 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months+Actual Sales Next 3 Months) (column 1); mean squared error (MSE) calculated as the square of the arc-percent forecast errors (column 2); win rate percentage, where winners are firms whose forecasts were within 10% of the realized value (column 3); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment (column 4); and over-precision, the probability to have the actual 3-months outcome outside of the predicted best and worst scenarios (column 5). The dependent variable mean is reported for the untreated control group. All regressions include wave-by-round fixed effects, and standard errors are clustered at the firm level. * p<0.1, ** p<0.05, *** p<0.01.

Figure 5. : The dashboard intervention compressed forecast errors



Note: In the upper-panel, we show quantile regressions of the impact of the dashboard experiment (a dummy taking value 1 if treated, 0 otherwise) on the Forecast Error, calculated using 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months). Effects shown from the 5th-95th percentile. Point estimates are solid point and the 95% confidence intervals are the bars. All regressions include wave by round fixed effects. Errors clustered at the firm level. Sample of 9,994 responses from waves 5-10. In the bottom panel, we show the raw distributions of forecast errors for dashboard treated and untreated firms, on the same sample.

B. *Attention: Randomizing Forecast Accuracy Rewards*

We also tested randomizing the reward for forecast accuracy. Perhaps inaccuracy was caused by firms failing to take our survey seriously, or the reward amounts were not high enough to elicit sufficient effort. The rewards randomization occurred in waves 5, 7 and 9 so the payments should have been salient, since most firms by the later rounds will have received multiple prior payments.

The experiment randomized 50 percent of respondents into a higher reward experiment. As background in waves 1 to 4 survey respondents were given a $25 Amazon gift card if their 3-month prediction was within 10 percent of their actual revenue. This reward was provided after the end of the quarter, prior to the next survey wave, to maximize salience. This randomization was independent by wave, and independent of all other experiments. Firms in the reward randomization had a 25 percent chance of having no reward at all. A further random 25 percent stayed with the base amount of $25, and the remaining 50 percent were randomly placed in eight equal sized groups across the values of $50, $100, $150, $200, $250, $300, $350, and $400. These rewards represent large material payments for correctly forecasting next quarter's sales revenue, particularly as they are presumably untaxed.[32] The maximum award of $400 is 16-times higher than the previous baseline of $25, creating a high spread of potential rewards.

We first show in Figure 6 upper panel that increasing the survey rewards led to a large increase in response time on the forecast question. The non-reward group took on average 50 seconds to respond to the survey while the groups with rewards of $300 or greater took about 110 seconds. This additional 60 seconds is a lot of time for a survey question, presumably reflecting extra time spent reading the question, analyzing any data, considering external factors that may influence sales, and potentially consulting other employees. Figure 6 bottom-panel shows one impact of the rewards is to reduce the average bias. Interestingly, even a small incentive of about $25–$50 seems to increase the time forecasting and reduce the bias by more than 50 percent. For rewards above $200 there is almost no mean bias. Both the time and mean bias impact highlight the importance of incentivized surveys – perhaps poor baseline forecast inaccuracy reflects respondees rushing surveys, and not being financially 'penalized' for unrealistic forecasts.

To investigate this further Table 6 analyzes the rewards randomization. Panel (A) column (1) shows the randomized incentives increased total survey time by 6.3 seconds per $100 of additional reward. In column (2) we see the randomization increased the time taken to complete the forecasting question by 14.5 seconds per $100 reward. In column (3) we see the randomization had no impact

[32]We chose the maximum reward of $400 to ensure we stayed below the $600 annual threshold for reporting income to the IRS. With a $400 reward the most any survey respondent could earn in a calendar year would be $500: one starting ($50) and one follow up ($25) fixed payments totaling $75, one regular win of $25 each, and one win of up to $400 (noting we randomized rewards only in rounds 5, 7 and 9). In practice, no respondent randomized into the $400 rewards correctly forecasted twice consecutively, so even this $500 bound was never met.

on the prior quarter's revenue reporting (effectively a placebo test). In column (4) we see firms also provided less rounded responses when offered higher rewards, with every additional $100 of reward reducing the share of trailing "0" values by 1.9 percentage points.

In Table 6 Panel (B) column (1) we see the rewards intervention cut the mean error (bias) by 1.9 points per $100. However, the reward intervention had no impact on mean squared error (noise). Finally, in column (3) we see each $100 of reward increased the win-rate by 0.9 percentage points. Therefore, a full $400 reward increased this by about 3.6 percentage points, from a 16.1 percent base. In columns (4) and (5) we see the rewards, however, had no impact on error predictability or over-precision.

In Table 6 Panel C we see there are no dynamic impacts of the rewards intervention. Despite inducing firms to substantially increase their forecasting time and accuracy in the current rounds, there is no persistent impact on accuracy.

In Figure 7 upper-panel we estimate quantile regressions for the impact of $100 of additional reward for percentiles from 5 percent to 95 percent and see uniform reduction in the forecast level. This suggests that, unlike the dashboard experiment, the whole distribution of observed biases has shifted down.[33] In the lower panel we see this shifts the distribution of forecast errors to the left, but does little to reduce total (absolute) forecast error. This is because inaccuracy is mainly driven by idiosyncratic bias and noise.

Overall, these results suggest that providing stronger financial rewards for forecast accuracy induces firms to spend substantially more time making their forecasts. This leads them to cut their average over-optimism. One explanation is motivated reasoning, with optimism helpful for motivating managers, employees and investors, with a material financial reward tamping this down (e.g. Kunda, 1990; Bénabou and Tirole, 2002). This is akin to the MBA saying "*Believe it to achieve it*". One interpretation of this is that managers may be aware they are over-confident and could shift down their expectations when making important decisions for which over-confidence is costly. However, since mean optimism bias accounts for only a small share of the total forecast error the overall accuracy does not improve by much.

For full transparency, we also analyze the effects of these first two successful treatments on firm performance. In Table A9, we show the relationship between forecasting accuracy, instrumented by the dashboard and rewards treatments and firm performance. We find that, albeit smaller in magnitude than the correlation, all the point coefficients have the ex-ante anticipated sign: higher accuracy is associated with higher productivity, growth, sales and survival. Nonetheless, the

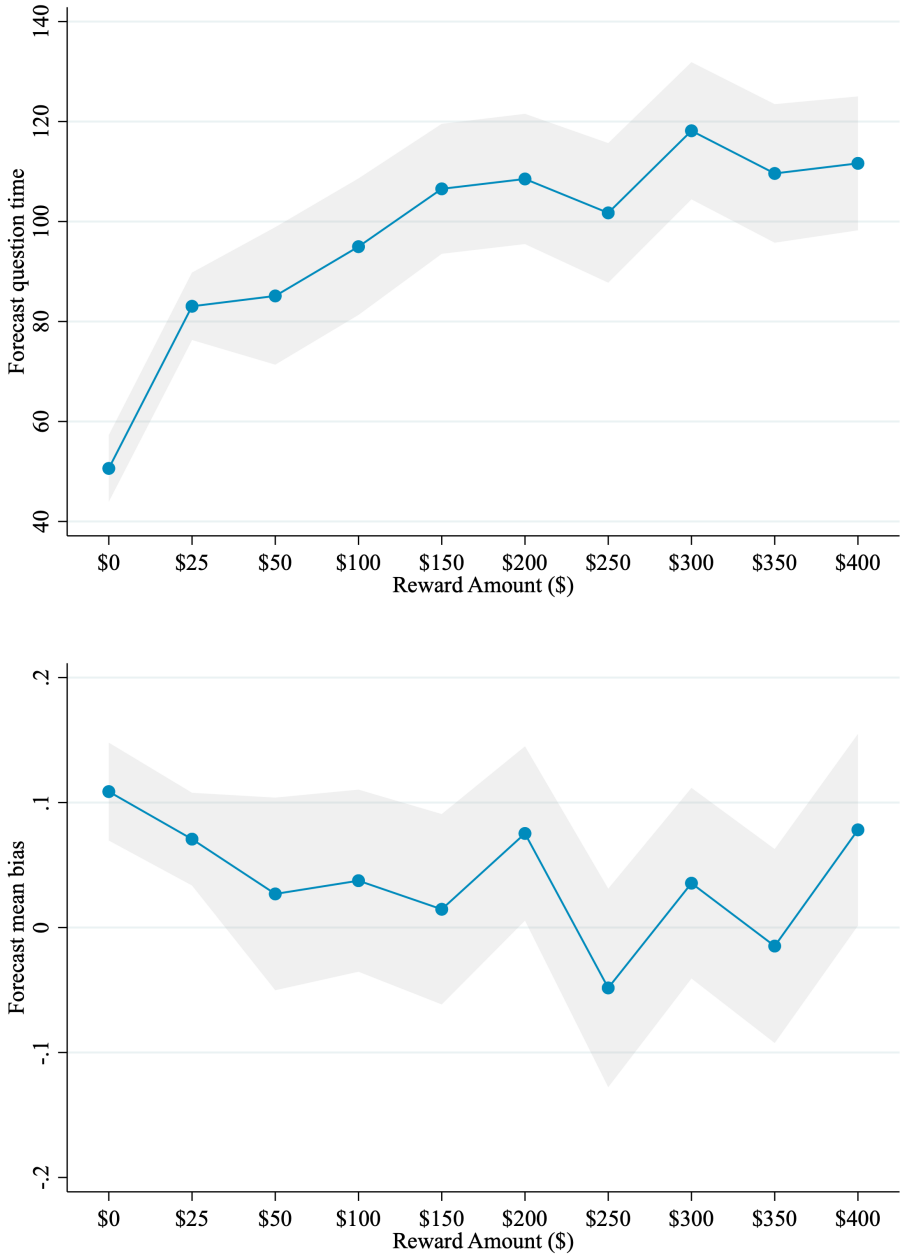
[33]The degree to which higher financial rewards decrease firms' over-optimism may be especially useful in models of rational inattention, where firms cannot process all of their available information, but are attentive to circumstances where the stakes are increased (e.g. Maćkowiak, Matějka and Wiederholt, 2023).

Table 6—: Reward Treatment (per $100)

Panel A: Forecast Diligence					
	(1) Survey Time (seconds)	(2) Prediction Time (seconds)	(3) Reporting Error^2	(4) Rounding (% digits)	(5)
Reward ('00 $)$_{i,t}$	6.321	14.510***	0.007	-1.853***	
	(4.686)	(1.422)	(0.009)	(0.216)	
N	6260	3357	6191	6260	
Dependent Mean	1191	50.6	0.344	62.3	
Panel B: Outcomes					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Reward ('00 $)$_{i,t}$	-0.019***	-0.009	0.009**	0.010	0.005
	(0.007)	(0.010)	(0.004)	(0.019)	(0.005)
N	6260	6260	6260	2668	6150
Dep. Mean	0.109	0.610	0.161	0.226	0.360
Panel C: Persistence					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Reward ('00 $)$_{i,t-1}$	-0.005	-0.021	-0.005	-0.027	0.001
	(0.010)	(0.014)	(0.006)	(0.024)	(0.007)
N	2926	2926	2926	1702	2882
Dependent Mean	0.077	0.527	0.211	0.301	0.406

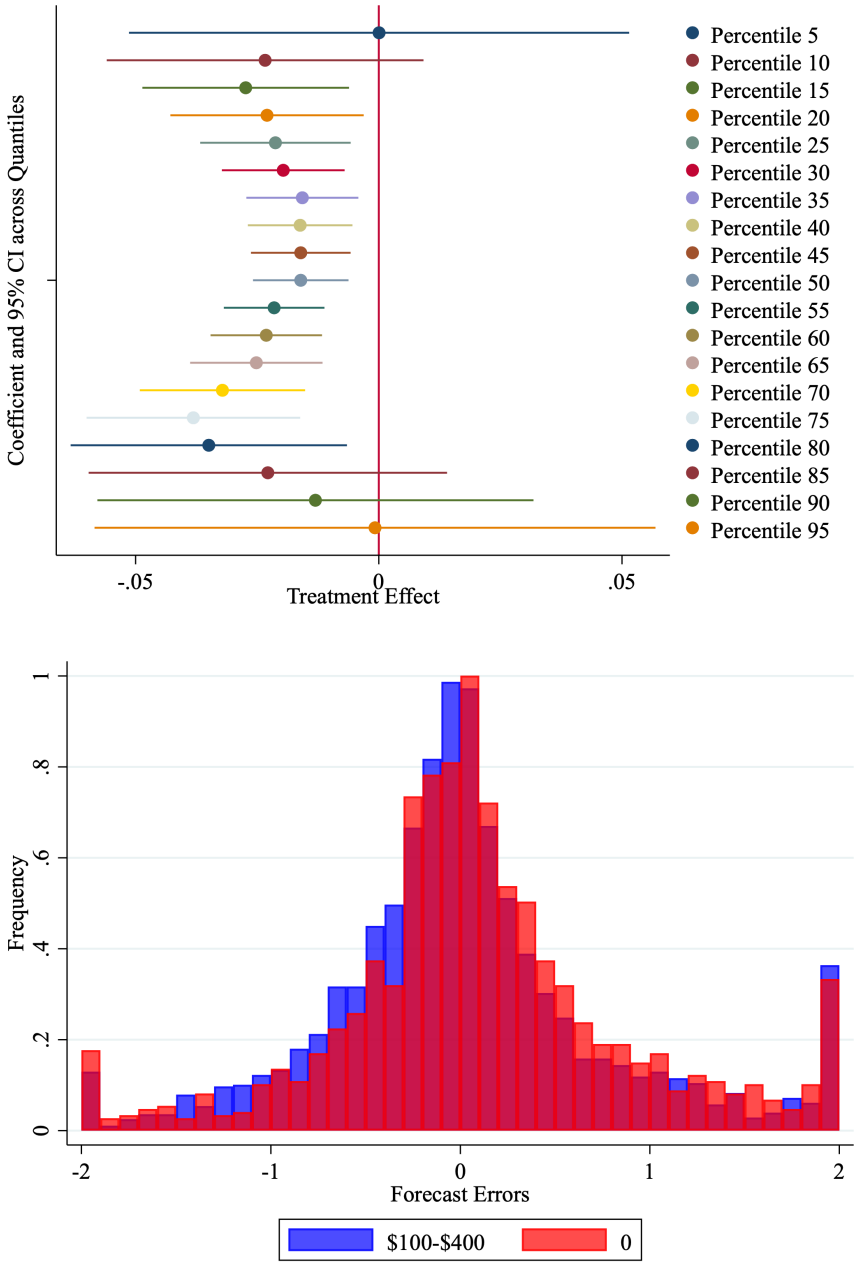
Notes: This table reports results of the reward treatment on survey and forecasting diligence (in Panel A), main forecasting outcomes (in Panel B), and persistence (in Panel C). The independent variable is the reward amount in hundreds of dollars (ranging from 0 to 4, representing $0 to $400) either this wave (Panels A and B) or last wave (Panel C).In Panel A, the dependent variables are: survey time, the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); prediction time, the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 10 minutes (column 2); historic sales squared error, calculated as [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 3); and rounding percentage of digits, the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panels B and C: forecast error (bias) which is calculated as 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months+Actual Sales Next 3 Months) (column 1); mean squared error (MSE) calculated as the square of the arc-percent forecast errors (column 2); win rate percentage, where winners are firms whose forecasts were within 10% of the realized value (column 3); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment (column 4); and over-precision, the probability to have the actual 3-months outcome outside of the predicted best and worst scenarios (column 5). The dependent variable mean is reported for the untreated control group. All regressions include wave-by-round fixed effects, and standard errors are clustered at the firm level. * p<0.1, ** p<0.05, *** p<0.01.

Figure 6. : Rewards increased survey time and decreased mean forecast bias



Note: These graphs show regression results of the timing spent predicting the incentivized 3-months sales question and the bias measured as the (arc-percentage), calculated as 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months). The former was winsorized at 10 minutes due to pauses during responding. There are 3,357 responses coming from 2,849 firms in the upper-panel, and 6,260 responses coming from 4,123 firms in the bottom-panel. Both figures plot the coefficients of a regression on the dummies of the rewards incentives, for the periods 5, 7 and 9 (when we ran the rewards randomization) controlling for full time fixed effects.

Figure 7. : The rewards intervention reduced the mean forecast error



Note: In the upper-panel side, we show regressions showing the impact of the reward experiment (in $'00s) on the Forecast Error, calculated using 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months). Effects shown from the 5th-95th percentile. Point estimates are solid point and the 95% confidence intervals are the bars. All regressions include wave by round fixed effects. Errors clustered at the firm level. Sample of 6,260 responses from waves 5, 7 and 9. In the bottom panel, we show the distributions of forecast errors for rewards treated (>=$100 reward) and untreated (0$ rewards) firms. Results are from a sample of 4,245 for waves 5,7, and 9 where the rewards RCT was used.

first stage F-statistics (between 6-7) show, unsurprisingly, that we have a weak instrument: the experiments only had a moderate impact on accuracy. Therefore, these results should be interpreted with caution.

C. Skills: Randomizing Forecast Training Exercises

After analyzing the results from waves 5 and 6 showing the moderate impacts of the data intervention and increased rewards, we decided to introduce a training experiment. The rationale was that perhaps the low success rate for firms in forecasting next quarter's revenue was due to limited experience in making forecasts, or issues understanding the question. Our hypothesis was perhaps if firms had experience of repeatedly making forecasts on a similar basis to the survey question, they would provide more accurate predictions.

So, in waves 7, 9 and 10 we randomized a forecast training exercise.[34] Randomization was independent across waves and independent of other interventions.

The training exercise asked firms to make forecasts for a series of up to ten hypothetical businesses. Respondents were first presented with a graph showing the historical revenue for a hypothetical firm and then were asked to predict the revenue for the firm in the next quarter (Figure A11 left panel). They were then shown a suggested forecast we created (e.g. also in Figure A11, right panel), which placed approximately 1/2 weight on the last quarter revenue and 1/6 on the prior three quarters revenues, based on a simple AR(4) best-fit regression of sales revenues for our sample frame.[35] This process was repeated for up to ten hypothetical firm vignettes, with the vignette ordering randomized by firm.

Before discussing the impact of the forecasting exercise, we present results validating this training exercise. We start by examining the progress of firms forecasting accuracy over the course of the randomly ordered vignettes. We found forecast errors appeared to roughly half from the first to last vignette as indicated in Figure 8 bottom-right panel. The variance of responses, the absolute distance from our simple "econometric" forecast, and the absolute distance from the mean forecast all fell by almost 50 percent over the course of the training. All of this suggests that managers were making material improvements in this forecasting exercise, which was designed to simulate forecasting next quarter's revenue for their own firms.

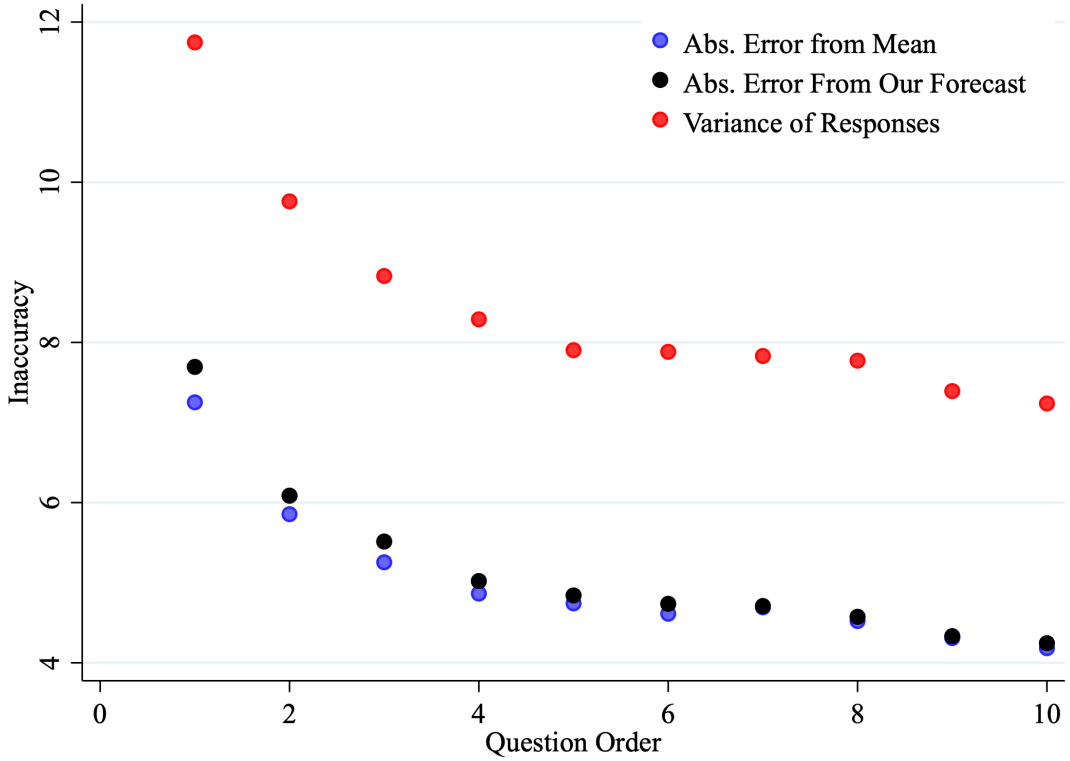
Moreover, firms' performance on this forecast training exercise was correlated with their prior

[34]The training randomization had three arms – one arm had the training right before making their forecast (treatment), a second arm had it right after making their forecast and a third arm had no training (control). The rationale for having three arms was that perhaps this training will not offer only immediate impacts on forecast accuracy, but also improvements in next waves that will be captured by both those seeing the exercises before forecasting and those seeing them after

[35]We regressed log(current) revenues on several lags of log revenues. Lags up to 4 were statistically significant, with weights very close to 0.5 on the first lag and 1/6 on the next three lags. Results were similar with or without industry and time fixed effects.

performance on actual revenue forecasts in two ways (details in Appendix Figure A7). First, firms' bias in prior forecasts was correlated with their bias in the forecasting exercise, with optimistic and pessimistic firms showing directionally similar biases on the training exercise. Second, firms forecast accuracy in their prior forecasts was correlated with the accuracy in the training exercise. So, firms with high accuracy in previously forecasting their own revenue had higher accuracy in the training exercise.

Figure 8. : Forecast training accuracy increases over time



Note: This figure shows results of three different ways of assessing accuracy for the set of firms randomized in the training treatment. We show the absolute error from mean, absolute error from our forecast and variance of responses, plotted against the randomized question order (N=27,412). Firms were first presented with a graph showing the historical revenue for a hypothetical firm and then were asked to predict the revenue for the firm in the next quarter (Figure A11 left panel). They were then shown a suggested forecast we created (e.g. also in Figure A11, right panel) based on a simple AR(4) best-fit regression of sales revenues for our sample frame. This shows that training almost halved forecast errors in the exercises between the first question and the last question–firms learnt from this exercise.

Despite this, however, as we see in Table 7 this training exercise had no impact on current or future forecasting accuracy. Table 7 first presents the survey and forecast diligence metrics. In column (1) we see firms randomized to the forecasting exercise took 45 seconds longer to complete the survey. In column (2) we see the training increased the prediction response time by 2.9 seconds,

Table 7—: The Training Treatment

Panel A: Forecast Diligence					
	(1) Survey Time (seconds)	(2) Prediction Time (seconds)	(3) Reporting Error^2	(4) Rounding (% digits)	(5)
Training$_{i,t}$	45.125***	2.875	0.059**	-0.634	
	(13.165)	(2.899)	(0.029)	(0.621)	
N	5703	4565	5586	5703	
Dependent Mean	1215	75.4	0.390	57.5	
Panel B: Outcomes					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Training$_{i,t}$	-0.009	0.060*	-0.014	0.072	0.013
	(0.023)	(0.032)	(0.011)	(0.064)	(0.014)
N	5703	5703	5703	2494	5586
Dep. Mean	0.096	0.630	0.180	0.261	0.387
Panel C: Persistence					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Training$_{i,t-1}$	-0.020	-0.014	0.013	0.025	0.016
	(0.036)	(0.049)	(0.019)	(0.093)	(0.024)
N	1931	1931	1931	1024	1896
Dependent Mean	0.125	0.591	0.203	0.300	0.426

Notes: This table reports results of the training treatment on survey and forecasting diligence (in Panel A), main forecasting outcomes (in Panel B), and persistence (in Panel C). The independent variable is a dummy taking the value 1 if a firm is treated by the training, and 0 otherwise either this wave (Panels A and B) or last wave (Panel C). In Panel A, the dependent variables are: survey time, the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); prediction time, the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 10 minutes (column 2); historic sales squared error, calculated as [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 3); and rounding percentage of digits, the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panels B and C: forecast error (bias) which is calculated as 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months+Actual Sales Next 3 Months) (column 1); mean squared error (MSE) calculated as the square of the arc-percent forecast errors (column 2); win rate percentage, where winners are firms whose forecasts were within 10% of the realized value (column 3); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment (column 4); and over-precision, the probability to have the actual 3-months outcome outside of the predicted best and worst scenarios (column 5). The dependent variable mean is reported for the untreated control group. All regressions include wave-by-round fixed effects, and standard errors are clustered at the firm level. * p<0.1, ** p<0.05, *** p<0.01.

although this was insignificant. In column (3) we see the forecasting exercise significantly reduced the accuracy of their last quarter's data reporting, perhaps because this training induced some survey fatigue. In column (4) we see it had no material impact on rounding, albeit with some suggestive negative impact.

In Table 7 panel B we examine the impact of the forecast training of their forecast accuracy. In columns (1) to (5) we see no significant or material impact on any current metrics. In Table 7 panel C we also see no dynamic impacts. Collectively, these results suggest that while forecast training improved firms' forecasting skills in the training exercise, it did not improve their forecasting skills for their own quarterly revenue. This appears to rule out that the firm's poor forecasting was primarily due to a lack of experience in quarterly revenue forecasting.

D. Contingent Thinking: Randomizing Consideration of Extreme Scenarios

After the null (or small improvements) from the interventions in waves 5 to 7, in wave 8 we tried one final treatment, which we repeated in wave 10. This asked firms to consider a wider set of scenarios before they make their forecasts. The literature around "Superforecasters" suggests the most successful forecasters consider a wide range of possible scenarios before making forecasts (e.g. Tetlock and Gardner, 2015; Haran and Moore, 2014). Although this literature focuses on experts forecasting various geo-political events, we thought perhaps a similar exercise could help firms. Clearly, firms face a number of uncertainties outside of their control such as macro cyclicality, movements in tax and interest rates, policy changes, or unexpected shifts in consumer demand. Perhaps the issue firms had with forecasting was they did not adequately consider the full range of scenarios, or their probabilities of occurring, consistent with their over-precision in Figure 4.

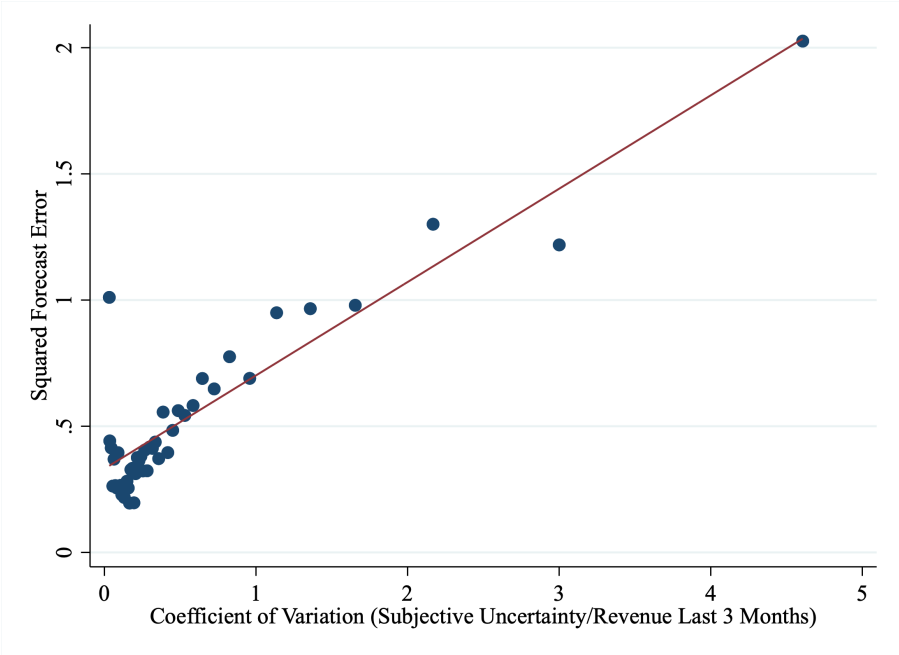
So, in waves 8 and 10 firms were randomized into the scenarios training experiment. This randomization was independent across waves and independent of other experiments. The randomization involved moving the subjective probability distribution question (shown in Appendix Figure A12), which asked firms about their lowest, low, middle, high and highest outcomes and the probabilities of these occurring from after the revenue forecasting question to immediately before it. The idea was that by moving this question immediately prior to the revenue forecasting question firms may make a more informed forecast about the most likely outcome. By asking them to consider scenarios from their lowest to higher outcomes it may reduce forecast tail errors.[36]

Before examining the results of the intervention, it is worth highlighting that firms' responses to these scenario questions appear reasonable. Figure 9 plots on the x-axis the implied subjective

[36]Technically, for an optimizing firm, the best answer to our incentivized 3-months forecasting question is to give something close to the mode of their sales distribution. Nonetheless, analyzing the data on all firms (see for example Figure 4) shows that the mode, mean and median of the sales distributions are quite close (overlapping) in our sample.

uncertainty from scenario questions versus the 3-month later realized squared forecast error on the y-axis. We see a highly significant relationship (t-stat of 25.5). Firms that provide a wide range of future sales scenarios have far larger subsequent 3-month forecast errors. So, on average firms appear aware of the uncertainty surrounding their own revenue forecast accuracy.

Figure 9. : Scenarios subjective uncertainty vs forecast error



Note: Binscatter (N=25) between squared 3-month forecasting error and the coefficient of variation, calculated from the subjective distribution from the 3-point or 5-point outcome probability forecast. The coefficient of variation is the subjective uncertainty measure, i.e. the standard deviation of the total variance from each term (squared deviations weighted by probabilities) divided by the revenue in the prior 3 months (when available). The coefficient of variation was winsorized at 5th and 95th percentile. Distribution observed in periods 5-10 for 9,819 responses of 4,266 firms. The linear unconditional relationship between these measures gives a t-statistic of 25.54 (with standard errors clustered at the firm level).

Table 8 presents the results of the scenarios intervention. We see virtually no material impact on firm forecasting. In panel (A), we evaluate the survey diligence metrics. We see this exercise has a negative impact on overall survey duration in column (1), decreasing it by 53.4 seconds. More than half of this decrease comes from reducing the time taken in the prediction exercise (shown in column 2), presumably because firms have undertaken a similar exercise to predict five scenarios immediately prior. The scenario treatment has no impact on last quarters reporting error in column (3). In column (4), we see it has no impact on the forecast rounding.

In Table 8 panel (B) we find across all five metrics – bias, noise, win-rate, error-predictability and over-precision– no impact of this exercise. So, having firms consider scenarios before making their forecasts does not appear to have any positive or negative impact on forecast accuracy or bias.

Table 8—: The Scenarios Treatment

Panel A: Forecast Diligence					
	(1) Survey Time (seconds)	(2) Prediction Time (seconds)	(3) Reporting Error^2	(4) Rounding (% digits)	(5)
Scenarios$_{i,t}$	-53.423***	-27.971***	0.065	-0.461	
	(19.326)	(3.011)	(0.046)	(0.903)	
N	2480	2480	2404	2480	
Dependent Mean	1232	66.8	0.457	56.1	
Panel B: Outcomes					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Scenarios$_{i,t}$	0.022	0.005	-0.006	0.007	0.023
	(0.034)	(0.050)	(0.016)	(0.066)	(0.020)
N	2480	2480	2480	1931	2422
Dep. Mean	0.175	0.759	0.198	0.256	0.435
Panel C: Persistence					
	Bias: Forecast Error	Noise: MSE	Accuracy: Win Rate	Error Predictability	Over-Precision
Scenarios$_{i,t-1}$	-0.026	-0.029	0.011	0.051	-0.009
	(0.051)	(0.068)	(0.029)	(0.118)	(0.035)
N	795	795	795	711	788
Dependent Mean	0.019	0.519	0.199	0.072	0.417

Notes: This table reports results of the scenarios treatment on survey and forecasting diligence (in Panel A), main forecasting outcomes (in Panel B), and persistence (in Panel C). The independent variable is a dummy taking the value 1 if a firm is treated by the scenarios, and 0 otherwise either this wave (Panels A and B) or last wave (Panel C). In Panel A, the dependent variables are: survey time, the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); prediction time, the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 10 minutes (column 2); historic sales squared error, calculated as [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 3); and rounding percentage of digits, the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panels B and C: forecast error (bias) which is calculated as 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months+Actual Sales Next 3 Months) (column 1); mean squared error (MSE) calculated as the square of the arc-percent forecast errors (column 2); win rate percentage, where winners are firms whose forecasts were within 10% of the realized value (column 3); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment (column 4); and over-precision, the probability to have the actual 3-months outcome outside of the predicted best and worst scenarios (column 5). The dependent variable mean is reported for the untreated control group. All regressions include wave-by-round fixed effects, and standard errors are clustered at the firm level. * p<0.1, ** p<0.05, *** p<0.01.

In Table 8 panel (C) we also see this scenario treatment also has no dynamic impact on future forecasting performance.

The null effect of scenario evaluation in this randomization appears to rule out firms making poor forecasts due to challenges of considering a wider set of possible outcomes. This treatment gave them practice making forecasts in advance of the incentivized question, including for more extreme outcomes, but with no impact on accuracy.

E. Robustness

In Tables A6 to Table A9 we run a number of robustness checks on our experiment results. The first check is to test whether the treatment effects are different on the sample of users with more than half of their revenue in the last quarter coming from FinTech (a sample which averages 82% of their revenue coming from FinTech). The second check is to exclude all firms below $100,000/year in income. Third, we introduce firm fixed effects to account for any time-invariant heterogeneity in firm characteristics. Fourth, we introduce the attrition weights, accounting for differential drop-out from the sample in our survey.

The results in Table A6 show the heterogeneity of the dashboard intervention treatment effects. Overall, we see the results are robust to all of the conditions above. In terms of effects on win rates these range from 1.1 to 3.8 percent (with a baseline of 2 percent). For over-precision, they range from 2.6 to 5 percent (with a baseline of 3.3 percent). The reduction in noise ranges from 3.1 to 12.2 percent (with a baseline of 9.7 percent). The moderate bias reduction results are the only ones that seem to be affected by these sample conditions, because as noted earlier they seem to be driven by smaller firms. The error predictability show no reduction throughout these sample cuts.

The results in Table A8 show the heterogeneity of the rewards intervention treatment effects. Overall, we see the results are robust to all of the conditions above. In terms of effects on win rates these range from 0.8 to 2.4 percent for every $100 in rewards (with a baseline of 0.9 percent). The reduction in bias ranges from 0.9 to 2.7 percent (with a baseline of 1.9 percent). These seem to be highest in the sample with the higher percentage of income on FinTech. The noise, error predictability and over-precision show no consistent reduction throughout these sample cuts.

The results in Table A7 and Table A9 show that no particular subsample showed any improvements in forecasting due to our training and scenarios interventions. The sporadic coefficients that are statistically significant at conventional levels seem to be caused by statistical noise and multiple hypothesis testing.

So, in summary the four experiments provide support for two root causes of forecasting biases. First, firms appear excessively confident in their forecasting ability and so underuse easily avail-

able prior information. This leads them to display over-precision in their subjective forecasts and consistently under-use data (like the dashboard) that would improve their forecast accuracy. Even when firms are randomized into the dashboard, which raises their current win-rates, they do not continue to use the dashboard, suggesting this over-confidence is persistent. Second, there is evidence for some pliable optimism bias, in that firms are over-optimistic on average but can temper this when provided with higher rewards. One explanation is motivated reasoning, in that firms find it useful to maintain an over-optimistic forecast but can temporarily suppress this when the stakes of their predictions are higher (such as when they are more highly rewarded for accuracy). This also suggests that the levels of bias found in prior studies (including those of authors of this paper, e.g. Altig et al., 2020) may be over-stated with no incentives for accuracy.

V. Importance of Our Results

In the previous sections, we have seen that forecasting accuracy is strongly correlated with performance, firms believe forecasting is important and despite that, forecasting issues among firms, including average bias, idionsycratic bias, predictable errors and over-precision are persistent and hard to fix. In this section we explore two reasons why, despite limited intervention impact, our results are important. First, we assess to what extent the results could be predicted by economists, and second we do a simple quantification exercise.

A. Shifting Priors

A first question is to what extent are these results anticipated or surprising. We used the Social Studies Prediction Platform to collect predictions on our experiments. The Social Studies Prediction Platform (SSPP) surveys a panel of researchers - mostly economics PhD students, academics and professionals - about their predictions for research projects. These predictions help interpret results, mitigate bias against null results and provide a baseline hypothesis for experimental interventions (see DellaVigna and Pope, 2018; DellaVigna, Pope and Vivalt, 2019).

Our study was available on the Social Science Prediction Platform for 2 months. A screenshot of the welcome message shown on SSPP is available in Appendix Figure A13. We attach more details on data collection, background, questions asked to respondees and elicitation in the Appendix D. Throughout this period, we obtained 53 complete predictions, which yielded two key results, outlined in Table 9.

First, in Table 9 Panel A, we see SSPP respondees expected some bias and over-precision, but about half the magnitude we found in the data. The SSPP predicted over-optimism of 9.6 percent compared to 16.3 percent in practice, and the share of outcomes outside the lowest/highest spread

of 18.6 percent compared to 39.4 percent in practice. Similarly, in Panel B we also see the SSPP predicted a far higher win-rate than participants achieved, forecasting a 46.8 percent win-rate compared to just 18 percent in practice.[37]

Finally, in Panel C we see that just like the authors of this study the SSPP respondees were over-optimistic on the impact of the randomized controlled trials on win rates. They expected that the data intervention would increase win-rates by 7.8 percent compared to 2.0 percent in practice. They expected the $400 incentives intervention would increase win-rates by 5.7 percent versus 3.5 percent in practice. And they predicted the most effective interventions would be the training and scenarios, which they predicted would increase win rates by 10.9 percent and 9.3 percent respectively, compared to no significant impacts in practice.

Table 9—: Social Science Prediction Platform Predictions vs Actual Results

	(1) Predicted	(2) Actual	(3) T-stat	(4) P-value
Panel A: Expected Biases				
Over-Optimism	9.6 (1.4)	16.3 (0.7)	-4.34	0.000
Over-Precision	18.6 (2.1)	39.4 (0.5)	-9.80	0.000
Panel B: Baseline Win Rate Expectation				
Win Rate (%)	46.8 (2.7)	18.0 (0.3)	10.67	0.000
Panel C: Treatment Effects on Win Rates				
Data	7.8 (1.1)	2.0 (0.8)	4.33	0.000
Incentives ($400)	5.7 (1.2)	3.5 (1.4)	1.19	0.236
Training	10.9 (1.5)	-1.4 (1.1)	6.68	0.000
Scenarios	9.3 (1.1)	-0.6 (1.6)	5.08	0.000

Notes: This table reports the difference between the predicted Social Science Prediction Platform biases and treatment effects and the actual results from our study. The first subpanel focuses on questions about the average win rates, over-optimism, and share of firms predicting sales outside of the (lowest, highest) interval. The second subpanel focuses on questions about the effectiveness of our treatments. Columns (3) and (4) report the results from a 2-sample Welch t-test and the 2-tail p-value associated with it. We test the null hypothesis that the two means are equal $H_0 : \Theta_{\text{pred}} - \Theta_{\text{act}} = 0$ with a t-test using $t = \frac{\widehat{\Theta_{\text{pred}}} - \widehat{\Theta_{\text{act}}}}{\sqrt{SE^2_{\text{pred}} + SE^2_{\text{act}}}}$. Stars * represent significance at 10% level i.e. p-value of mean being 0 is <0.1 ** p-value is <0.05 *** p-value is <0.01.

One explanation for this over-optimism in the SSPP forecasts is respondents based their predictions on the prior economics literature. As seen in the literature review, this is mostly focused on large, often listed firms. Large firms often have professional forecasting teams, generating more

[37]We note that the SSPP vignette used a comparison of the $400 reward compared to the $25 accuracy bonus as that was the amount at baseline; the regression analysis in Table 6, and the "actual" figure in Table 9, instead normalise the reward variable so that $0 is the omitted category. This does not change any substantive result. For the same simplicity reason, SSPP respondents were asked about the standard percentage-error while the paper reports arc-percentage; at the realised magnitudes the two measures differ by only about 1.5 percentage points, far smaller than the 7 percentage points gap between SSPP beliefs and actual bias.

accurate forecasts. These professional forecasting units may also be expected to respond more strongly to improved data, incentives or training, suggesting perhaps a larger treatment impact of the interventions. The fact we find very different results highlights the value of collecting forecasting performance and experimental data from the full size distribution of firms.

B. Quantification of Economic Costs

To understand the economic significance of the forecasting issues documented in Section III, we build up an illustrative model of firm investment. The model allows us to quantify the potential costs of these biases compared to a rational expectations benchmark. We simulate three different scenarios: one is rational expectations, where we assume no biases on firm forecasting, one where these biases are in line with those found in the Social Science Prediction Platform survey results ("Expected Biases"), and one that is representative of the average biases found in our data ("Actual Biases").

Throughout, we will use the simplest modelling choices to highlight the three main consequences of poor forecasting: suboptimal input choice, welfare losses and misallocation.

We model a population of firms, indexed by i, where each firm produces price-normalized output (Y_{it}) using a single input, capital (K_{it}) in a Cobb-Douglas production function:

$$Y_{it} = pA_{it}K_{it}^{\alpha} \tag{7}$$

where A_{it} is a stochastic productivity shock, α is the capital share, and p is the output price.

Each firm faces stochastic productivity shocks, A_{it}, which follow a persistent, autoregressive process:

$$A_{i,t+1} = (1-\rho)\bar{A} + \rho A_{it} + \varepsilon_{it+1}, \quad \varepsilon_{it+1} \sim N(0, \sigma_{\varepsilon}^2) \tag{8}$$

where ρ is the persistence parameter and $\bar{A}$ is the long-run mean productivity, which will be normalized at 1.

Forecasting is important in this setting because adjusting the capital stock is costly. We assume a quadratic adjustment cost, which penalizes large changes in the capital stock. Each firm chooses its next-period capital stock, K_{it+1}, to maximize the expected present value of future profits. The problem for firm i is described by the Bellman equation:

$$V(K_{it}, A_{it}) = \max_{K_{i,t+1}} \left\{ pA_{it}K_{it}^{\alpha} - (r+\delta)K_{it} - C(K_{i,t+1}, K_{it}) + \frac{1}{1+r}E_t^F \left[V(K_{i,t+1}, A_{i,t+1}) \right] \right\} \tag{9}$$

where the total cost of investment, $C(K_{i,t+1}, K_{it})$ is of quadratic form:[38]

$$C(K_{i,t+1}, K_{it}) = \frac{\xi}{2} \frac{(K_{i,t+1} - (1-\delta)K_{it})^2}{K_{it}} \tag{10}$$

where r is the real interest rate, δ is the depreciation rate, ξ is the quadratic adjustment cost parameter, and E_t^F is the firm's subjective expectation operator, which may deviate from the true, rational expectation E_t. In other words the firm is choosing the next period capital, maximizing its revenue minus the user cost of capital (for simplicity, we assume capital is rented), minus the adjustment cost of investment, and adding the future discounted value (perceived) coming from this capital choice.

We simulate an economy with 5,000 firms operating on a quarterly frequency. All parameter values are reported in Table A10. The capital (only input) share is set to $\alpha = 0.3$. The interest rate is $r = 0.0259$ per quarter (approximately 10% annually), and the depreciation rate is $\delta = 0.0259$ per quarter. The true persistence parameter for productivity is $\rho = 0.9$. The standard deviation of productivity shocks, σ_ε, is 0.1. The quadratic adjustment cost parameter is $\xi = 0.5$.

To simulate the impact of the empirical forecasting biases documented in Section III, we compare the baseline rational expectations model to alternative specifications incorporating each of the key biases we identify. In the behavioral models, firms make decisions based on a perceived productivity state, A_{it}^P, and use a policy function derived from an incorrect belief about the data-generating process. We model three distinct departures from rational expectations that capture the systematic patterns in our data:

Over-optimism: Firms are systematically optimistic about their productivity. We model this as a firm-specific, time-invariant bias, b_i, added to their perception of productivity. For each firm i, this bias is drawn from a normal distribution for the actual biases, $b_i^A \sim N(0.163, 0.08^2)$. This specification reflects the average optimistic bias in our data of 16.3% and roughly the heterogeneity of firm level biases (as seen for example in the sample in Table 3). For the "expected" biases, we use the mean expected over-optimism from Table 9, so $b_i^E \sim N(0.096, 0.08^2)$, keeping the standard deviation the same.

Predictable Errors: We model predictable errors as a form of inattention, where firms do not always use the most current information. A firm's perception is based on a potentially lagged productivity state, $A_{i,t-l}$. Based on Table 4 results on the dynamic autocorrelation of the actual biases, the lag l, is chosen stochastically each period according to the following probabilities: $P^A(l = 0) = 0.50$, $P^A(l = 1) = 0.20$, $P^A(l = 2) = 0.10$, $P^A(l = 3) = 0.10$, and $P^A(l = 4) = 0.10$. A firm's

[38] Quadratic adjustment costs were used for simplicity. Note that the results are not dependent on the type of adjustment cost on the input parameter. Any type of adjustment costs introduces costs of systematic errors in forecasting.

perceived state in terms of "actual biases" is thus $A_{it}^A = A_{i,t-l} + b_i$, combining inattention with their optimistic bias. While in the Social Science Prediction Platform we have not asked about this as we couldn't have elicited it in a simple question, we assume that the expected biases are roughly half, in line with the other expectations, therefore $P^E(l = 0) = 0.75$, $P^E(l = 1) = 0.10$, $P^E(l = 2) = 0.05$, $P^E(l = 3) = 0.05$, and $P^E(l = 4) = 0.05$.

Over-precision: Firms are over-precise in their beliefs about future productivity. We model this as an overestimation of the persistence of the productivity process. While the true persistence parameter is $\rho = 0.9$, for the "actual biases" firms believe it to be $\rho^A = 0.94$, in line with Figure 4. For the "expected biases" firms believe the persistence parameter to be $\rho^E = 0.92$, which is in line with the expected over-precision share of the actual over-precision in Table 9. By believing the process is more persistent, firms act as if the current state is a stronger signal of future productivity than it actually is, leading them to solve for a policy function based on an incorrectly perceived law of motion.

Table 10 presents the impact of these forecasting biases on key moments of the simulated economy, reported as percentage deviations from the rational expectations baseline. The results highlight three main consequences of these biases: over-investment, reduced profitability, and increased misallocation.

First, the combination of optimism, inattention, and over-precision leads firms to substantially over-invest. In the "Actual Biases" model, the mean capital stock is 16.5% higher than the rational benchmark. This is driven primarily by over-optimism; believing future productivity will be higher incentivizes firms to build up a larger capital stock in anticipation. While this leads to higher average output (+4.7%), it comes at a cost.

Second, this over-investment is ultimately unprofitable. The "Actual Biases" model results in a 14% decrease in mean profits over the unit of input. Firms are systematically choosing an input level that is too high for the productivity they actually realize, leading to a net loss despite higher output. We note that these numbers are very similar to Thesmar (2025) (using the Ma et al. (2024) framework) that calculates a potential profit gain of around 11% (mostly coming from removing conditional/predictable errors) in our sample.

Third, the biases generate significant misallocation and volatility. We measure misallocation by the correlation between a firm's true productivity and its capital stock. This correlation falls by 2.6% in the "Actual Biases" model, indicating that capital is less effectively allocated to the most productive firms. This is a result of both predictable errors and idiosyncratic optimism. In terms of volatility, over-precision causes firms to over-react to shocks they perceive as highly persistent. However, this is counteracted by optimism and inattention, which tend to smooth investment

decisions. The net result is a moderate increase in the volatility of output and profits.

In sum, the forecasting errors we document empirically can generate economically meaningful costs, distorting firm-level investment, decreasing firm profits and distorting allocative efficiency. Importantly, as can be seen from the percentage deviation from rational expectations column, the true extent of these biases matter.

Table 10—: Impact of Forecasting Biases on Economic Outcomes

	Model Values			**% Deviation from RE**	
Variable	RE	Expected	Actual	Expected	Actual
Mean Output (Y)	0.079	0.081	0.083	+2.30%	+4.68%
Mean Capital (K)	0.455	0.491	0.530	+7.89%	+16.45%
Mean Investment Rate (I/K)	0.027	0.027	0.027	+0.56%	+0.74%
Mean Profit/Capital (π/K)	0.121	0.112	0.104	-7.15%	-14.13%
Volatility of Output	0.274	0.278	0.280	+1.60%	+2.29%
Volatility of Capital	0.147	0.171	0.180	+16.30%	+22.28%
Volatility of Profits	0.383	0.390	0.400	+1.73%	+4.39%
Correlation(A, K)	0.581	0.562	0.566	-3.23%	-2.55%

Notes: This table presents simulation results from the neoclassical investment model with quadratic adjustment costs, comparing rational expectations (RE) to behavioral specifications incorporating the forecasting biases documented in Section III. The model simulates 5,000 firms over 1,000 periods on a quarterly frequency, with the first 100 periods discarded as burn-in. "Expected Biases" incorporates forecasting biases in line with those predicted by the Social Science Prediction Platform survey: over-optimism mean = 0.096 (from Table 9), perceived persistence $\rho^E = 0.92$, and reduced inattention (75% use current information). "Actual Biases" incorporates the average biases found in our firm-level data: over-optimism mean = 0.163 (16.3% average optimistic bias), perceived persistence $\rho^A = 0.94$ (from Figure 4), and inattention probabilities from Table 4 (50% use current information). Mean Profit is $\pi = pAK^{\alpha} - (r+\delta)K - C(K', K)$. Volatility measures are standard deviations of log values. Complete parameter specifications are reported in Table A10.

VI. Conclusion

Firms' expectations are a critical aspect of economics. However, evidence for the rationality and biases of firms' sales expectations is limited. This is due to issues with the existing literature around sample representativeness and response accuracy. Sample representativeness is an issue because the prior literature focuses on large often publicly listed companies. These large firms typically have professional forecasters or teams, so may have different biases from the majority of firms which are smaller with individual owners typically making forecasts. Response and follow-up accuracy can also be an issue in unincentivized surveys.

To develop this literature we take two steps. First, we partner with a large payment-processing company to create a sample of firms with a size distribution similar to the universe of all US employee firms. Using this panel we run a 5 year, 10 wave panel survey of quarterly sales forecasts

for 6,594 firms. Second, to elicit accurate responses we provide a $50 initial base fee, $25 per round additional fees, plus accuracy fees of up to $400 per response. These forecasts are evaluated against firms' actual payments-processing revenue data, allowing a precise evaluation of their accuracy.

We found three main types of forecast issues. First, forecasts are biased, with a mean optimism bias of 16 percent. Importantly, this bias has a wide idiosyncratic variation, as over-optimism is highly heterogeneous across firms. Second, firms have predictable errors in that forecast errors are correlated with lagged forecast errors, but also other pre-determined variables such as lagged growth and sales. Third, firms are over-precise. Our survey elicited not only point-forecasts but also second-moment forecasts using 3-bin and 5-bin outcomes and probabilities. Firms provided subjective forecast variances that were about half of the actual forecast variances. Perhaps unsurprisingly as the previous literature focused on large firms, we find that economists generally underestimate how widespread these biases and inaccuracies are.

Having identified these forecasting biases the question is then what causes them? Perhaps firms are acting as rational, profit maximizing agents, with for example over-optimism a way to motivate employees and investors. Or perhaps biases are costly, arising from poor management practices or the inability of individual owners to make good forecasts. To investigate this, we also ran four randomized controlled trials.

First, we randomized firms' exposure to an easily-available data-dashboard to evaluate the extent to which reviewing historical sales data improves forecasting. Second, we randomized accuracy rewards, ranging from $0 to $400, to evaluate to what extent greater incentives for accuracy improves forecasting. Third, we randomized forecast training, providing up to 10 vignettes of a forecasting exercise for a hypothetical firm to see if forecasting improves with experience. Finally, we randomized firms into evaluating scenarios in advance of making their forecasts, building on the "Superforecaster" literature.

We found data exposure through the dashboard reduced inaccuracy, bias and over-precision. Firms treated to the data dashboard took just 45 seconds more to complete the survey on average. This suggests even short data exposures can improve forecasting performance. Surprisingly, despite the dashboard being free to access and generating perhaps about $500 an hour of increased rewards, firms that were randomized into the dashboard in a current wave did not show improved forecast accuracy in future waves. This suggests firms are over-confident in their forecasting ability, under-appreciating the value of additional information.

The reward randomization reduced over-optimism bias. Firms randomized into the highest $400 reward bucket took more than twice as long to make their forecasts – implying greater care and consideration. For firms in the highest reward bucket this led to an almost full elimination of over-

optimism. This implies firms' over-optimism is pliable – firms are aware of it, and if adequately incentivized, can reduce their bias. Being over-optimistic may help self-motivate managers to work harder, but if provided with sufficiently generous rewards they can be more realistic.

Neither the forecast training nor the contingent thinking interventions had any impact on bias, error predictability or over-precision. As a result, they also had no impact on win rates. This suggests firm forecasting inaccuracy does not arise from a lack of forecast training or consideration of tail outcomes. We think this is an interesting result, especially as we can formally reject the economists' predictions that these would be the most effective interventions.

We think these results pave the way for future research. First, it highlights the importance of analyzing the full distribution of firm sizes. The management and operational practices of large publicly listed firms appear to be very different from the non-large firms. While large firms are important in aggregate trends, analyzing the full firm size distribution is valuable for economic modelling and policy. For example, young and small firms are an important source of job creation and innovation, and play a critical role in many IO, macro and trade models. Second, it highlights an important role for interventions to leverage data availability and analysis to improve firm forecasting. This is perhaps even more important as machine learning and AI tools offering personalized insights and predictions are rapidly advancing. Finally, embedding these forecasting behaviors into general equilibrium frameworks will identify broader macroeconomic implications and inform policy interventions aimed at improving resource allocation under uncertainty.

REFERENCES

Abel, A., and J. Eberly. 1996. "Optimal Investment with Costly Reversibility." *Review of Economic Studies*, 63: 581–593.

Afrouzi, Hassan, Spencer Y Kwon, Augustin Landier, Yueran Ma, and David Thesmar. 2023. "Overreaction in Expectations: Evidence and Theory*." *The Quarterly Journal of Economics*, 138(3): 1713–1764.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Brent Meyer, and Nicholas Parker. 2020. "Surveying Business Uncertainty." *Journal of Econometrics*.

Bachmann, Rüdiger, and Steffen Elstner. 2015. "Firm Optimism and Pessimism." *European Economic Review*, 79: 297–325.

Barrero, Jose Maria. 2022. "The micro and macro of managerial beliefs." *Journal of Financial Economics*.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics*, 117(3): 871–915.

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives*, 30(3): 141–164.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. 2013. "Managerial Miscalibration." *The Quarterly Journal of Economics*, 128(4): 1547–1584.

Bertola, Giuseppe, and Ricardo J. Caballero. 1994. "Irreversibility and aggregate investment." *The Review of Economic Studies*, 61(2): 223–246.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review*, 91: 67–72.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What Drives Differences in Management Practices?" *American Economic Review*, 109(5): 1648–1683.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review*, 110(9): 2748–82.

Boutros, Michael, Itzhak Ben-David, John R. Graham, Campbell R. Harvey, and John W. Payne. 2024. "The Persistence of Miscalibration." National Bureau of Economic Research NBER Working Paper 28010.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2018. "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico." *Journal of Political Economy*, 126(2): 635–687.

Cai, Jing, and Shing-Yi Wang. 2022. "Improving Management Through Worker Evaluations: Evidence from Auto Manufacturing." *The Quarterly Journal of Economics*, 137(4): 2459–2497.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8).

Coibion, Olivier, Gorodnichenko Y. Kumar S. 2018. "How Do Firms Form Their Expectations? New Survey Evidence." *American Economic Review*, 108: 2671–2713.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2020. "Inflation Expectations and Firm Decisions: New Causal Evidence." *The Quarterly Journal of Economics*, 135: 165–219.

Cooper, Russell W., and John C. Haltiwanger. 2006. "On the Nature of Capital Adjustment Costs." *The Review of Economic Studies*, 73(3): 611–633.

Davis, Steven J., and John Haltiwanger. 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation." *The Quarterly Journal of Economics*, 107: 819–863.

DellaVigna, Stefano, and Devin Pope. 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy*, 126: 2410–2456.

DellaVigna, Stefano, Devin Pope, and Eva Vivalt. 2019. "Predict Science to Improve Science." *Science*, 366(6464): 427–428.

Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment Under Uncertainty.* Princeton, NJ:Princeton University Press.

Farmer, Leland, Emi Nakamura, and Jon Steinsson. 2024. "Learning about the Long-Run." *Journal of Political Economy*, 132(10).

Federal Reserve Financial Services. 2024. "Federal Reserve Surveys: U.S. Businesses, Consumers Increasingly Adopt Faster and Instant Payment Services." *https://www.frbservices.org/news/press-releases/050624-faster-instant-payment-surveys.*

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein. 1977. "Knowing with Certainty: The Appropriateness of Extreme Confidence." *Journal of Experimental Psychology: Human Perception and Performance*, 3(4): 552–564.

Floyd, Eric, and John A. List. 2016. "Using Field Experiments in Accounting and Finance." SSRN Working Paper.

Gabaix, Xavier. 2023. "Behavioral Macroeconomics via Sparse Dynamic Programming." *Journal of the European Economic Association*, 21(6): 2327–2376.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer. 2016. "Expectations and Investment." *NBER Macroeconomics Annual*, 30: 379–431.

Gosnell, Greer K., John A. List, and Robert D. Metcalfe. 2020. "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains." *Journal of Political Economy*, 128(4): 1195–1233.

Graham, John R., Campbell R. Harvey, and Manju Puri. 2013. "Managerial Attitudes and Corporate Actions." *Journal of Financial Economics*, 109: 103–121.

Haran, Uriel, and Don Moore. 2014. "A Better Way to Forecast." *California Management Review*.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average Q: A Neoclassical Interpretation." *Econometrica*, 50: 213–224.

Huffman, David, Collin Raymond, and Julia Shvets. 2022. "Persistent Overconfidence and Biased Memory: Evidence from Managers." *American Economic Review*, 112(10): 3141–3175.

Isaacson, Walter. 2011. *Steve Jobs.* New York:Simon and Schuster.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5): 1449–1475.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* New York:Farrar, Straus and Giroux.

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. 2021. *Noise: A Flaw in Human Judgment.* Hachette UK.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin*, 108(3): 480–498.

Landier, Augustin, and David Thesmar. 2009. "Financial contracting with optimistic entrepreneurs." *The Review of Financial Studies*, 22(1): 117–150.

List, John A. 2011. "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off." *Journal of Economic Perspectives*, 25(3): 3–16.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational Inattention: A Review." *Journal of Economic Literature*, 61(1): 226–273.

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *Journal of Finance*, 60: 2661–2700.

Malmendier, Ulrike, and Geoffrey Tate. 2008. "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction." *Journal of Financial Economics*, 89: 20–43.

Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics*, 117: 1295–1328.

Manski, Charles F. 2018. "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise." *NBER Macroeconomics Annual*, 32: 411–471.

Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar. 2024. "A Quantitative Analysis of Distortions in Managerial Forecasts." National Bureau of Economic Research NBER Working Paper 26842. Revised version.

McKenzie, David. 2018. "Can business owners form accurate counterfactuals? Eliciting treatment and control beliefs about their outcomes in the alternative treatment status." *Journal of Business & Economic Statistics*, 36(4): 714–722.

McKenzie, David, and Dario Sansone. 2019. "Predicting Entrepreneurial Success is Hard: Evidence from a Business Plan Competition in Nigeria." *Journal of Development Economics*,

141: 102369.

Meyer, Brent, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Kevin Foster, Emil Mihaylov, and Aaron Jalca. 2025. "Survey of Business Uncertainty Monthly Report: September 2025." Federal Reserve Bank of Atlanta Monthly Report.

Moll, Benjamin. 2025. "The Trouble with Rational Expectations in Heterogeneous Agent Models: A Challenge for Macroeconomics."

Moore, Don A., and Paul J. Healy. 2008. "The trouble with overconfidence." *Psychological Review*, 115(2): 502–517.

Moskowitz, Toby, and Annette Vissing-Jorgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review*, 745–778.

Sharot, Tali, C. W. Korn, and R. J. Dolan. 2011. "How Unrealistic Optimism is Maintained in the Face of Reality." *Nature Neuroscience*, 14(11): 1475–1479.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics*, 50: 665–690.

Soll, Jack B., and Joshua Klayman. 2004. "Overconfidence in Interval Estimates." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2): 299–314.

Tanaka, Mari, Nicholas Bloom, Joel M. David, and Maiko Koga. 2019. "Firm Performance and Macro Forecast Accuracy." *Journal of Monetary Economics*.

Taubinsky, Dmitry, Luigi Butera, Matteo Saccarola, and Chen Lian. 2025. "Beliefs About the Economy Are Excessively Sensitive to Household-Level Shocks: Evidence from Linked Survey and Administrative Data." National Bureau of Economic Research NBER Working Paper 32664.

Tetlock, Philip E., and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction.* New York:Crown Publishers.

Thesmar, David. 2025. "Discussion of "Rationalizing Firm Forecasts" by Bloom, Codreanu, and Fletcher." *Prepared for the NBER Organizational Economics Working Group Fall 2025 Meeting*, Discussant commentary.

U.S. Census Bureau. 2021. "Management and Organizational Practices Survey (MOPS)." *https://www.census.gov/programs-surveys/mops.html*.

U.S. Census Bureau. 2022. "Business Dynamics Statistics Datasets 2022." *https://www.census.gov/data/datasets/time-series/econ/bds/bds-datasets.html*.

U.S. Census Bureau. 2023. "Nonemployer Statistics 2023." *https://www.census.gov/programs-surveys/nonemployer-statistics.html*.

Weinstein, Neil D. 1980. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology*, 39(5): 806–820.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review*, 110(2): 337–361.

Additional Tables and Figures

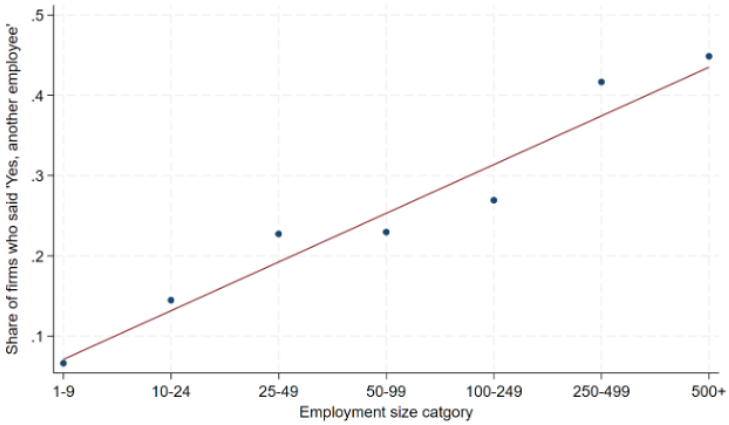


Figure A1. : Employment size bins and percentage of firms who say there's any other person (not CEO/CFO) whose job is to develop sales revenue forecasts (from Meyer et al. (2025))

Note: This figure is taken from the Survey of Business Uncertainty Monthly Report for September 2025 by Meyer et al. (2025). The survey, conducted by the Federal Reserve Bank of Atlanta from September 8-19, 2025, covers all U.S. states and major industry sectors, with a sample size of 1,139 respondents. These are responses to the question: "Is there anyone in your firm, including yourself, whose job it is to develop sales revenue forecasts?" The data shows a positive relationship between company size and the likelihood of responding "Yes, another employee" who is responsible for sales revenue forecasting.

Figure A2. : Employment size bins and aggregated average structured management scores (from U.S. Census Bureau (2021))

Note: This figure is based on data from the US Census Management and Organizational Practices survey 2021, n=36,000, available at https://www.census.gov/programssurveys/mops.html (U.S. Census Bureau, 2021). Management scores on a 0 to 1 scale across 16 questions, with a 0 denoting the least and 1 denoting the most structured management practices for monitoring, targets and incentives.

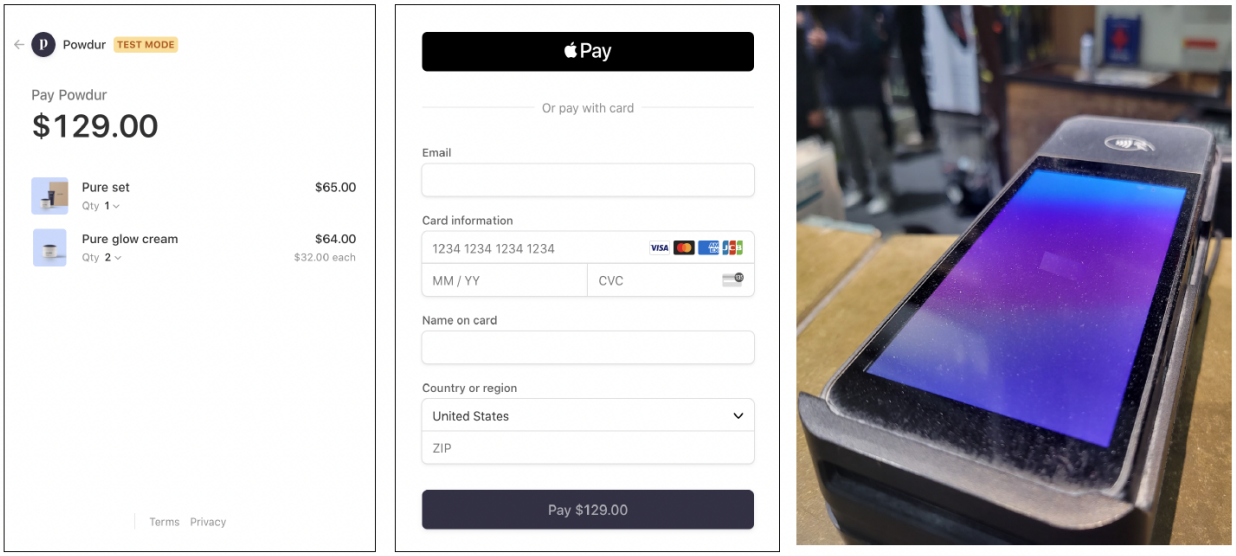


Figure A3. : Fintech Processes Payments Online and in Person

Note: This figure shows examples of the types of online (left two) and in-person (right) payment options used by the FinTech's customers. The name of the company has been covered or blurred for anonymity. The right photo was taken by a co-author in a local gym.

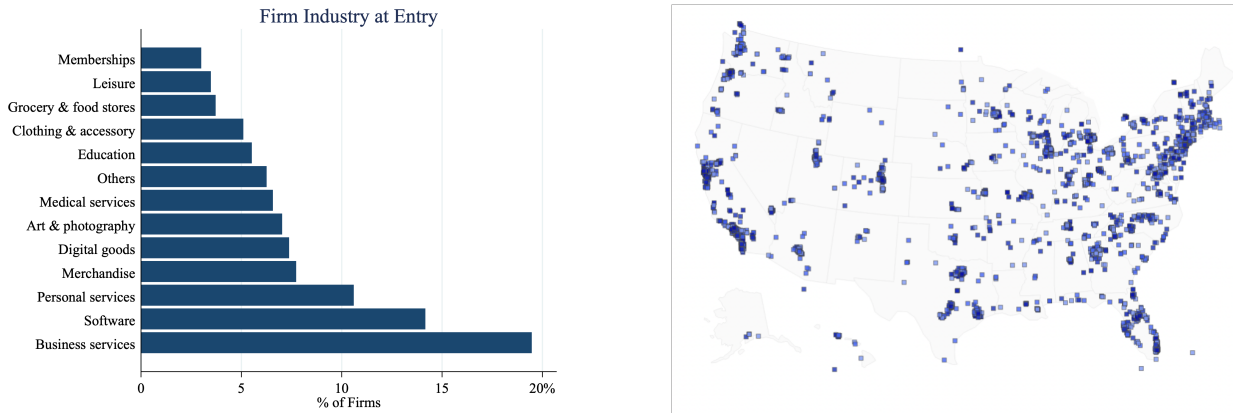


Figure A4. : Broad Industry and Region Spread of Firms

Note: On the left hand side, we show the histogram of 6,594 firms responding at least once during our study by industry, at entry, where each observation is equal weighted, while the right plot shows the location of a quasi-random sample of 5,800 firms we received answers from.

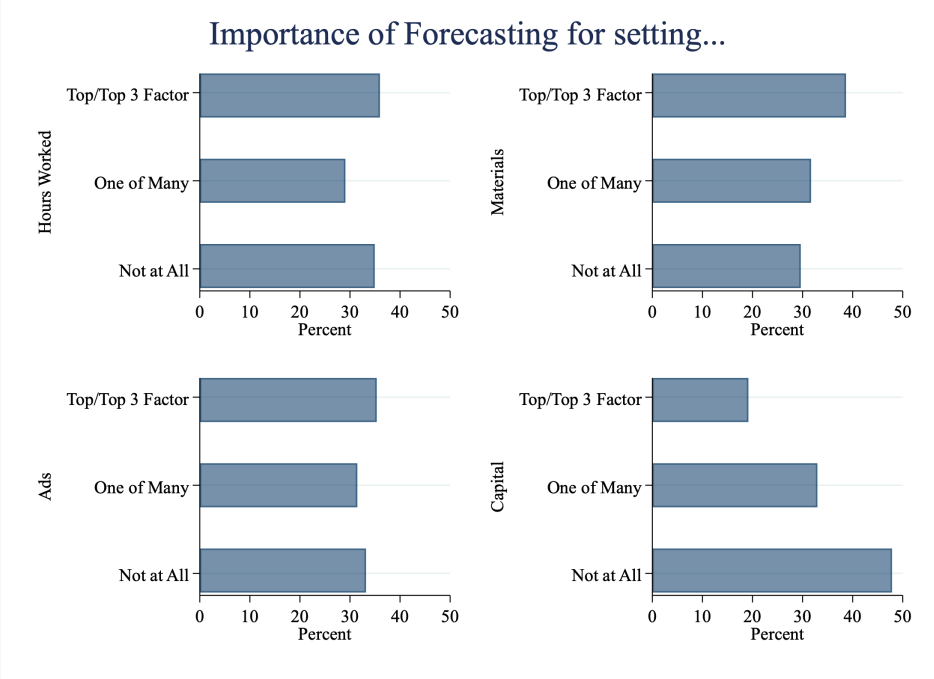


Figure A5. : Self-reported Importance of Forecasting

Note: This figure shows results from a question asking "How important is sales forecasting for deciding [e.g. hours worked]?".Results are all from wave 8, for 1,262 responses of individuals who evaluated the (self-assessed) usefulness of sales forecasting for their business. Results shown from top left to bottom right for the questions referring to the importance of forecasting setting hours worked, materials, advertising and capital.

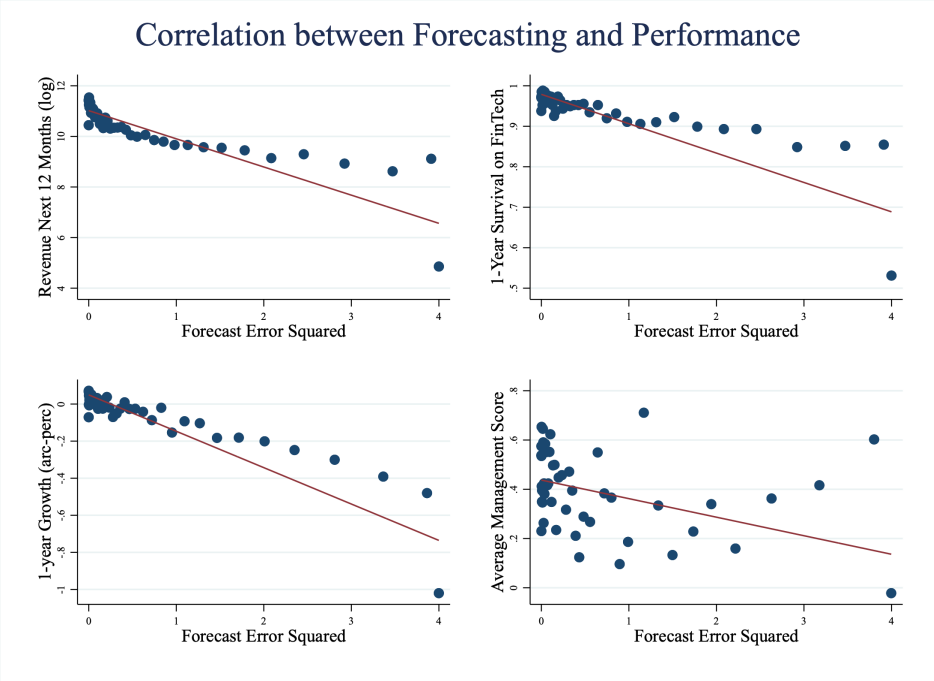


Figure A6. : Firm Performance and Forecasting

Note: This figure shows a panel of binscatters of the correlations between revenue of the firm in the next 12 months (log), survival (any positive income on FinTech) in the period 7-12 months after the survey, growth level (measured by 2*(Actual Sales Next 12 Months-Actual Sales Last 12 Months) / (Actual Sales Next 12 Months + Actual Sales Last 12 Months)) and average management score with the squared 3-month forecast error (arc-percentage). Results shown for the full sample when the necessary data were available. The linear unconditional relationship between these measures gives t-statistics of -76.83, -45.17, -49.39, and -3.63.

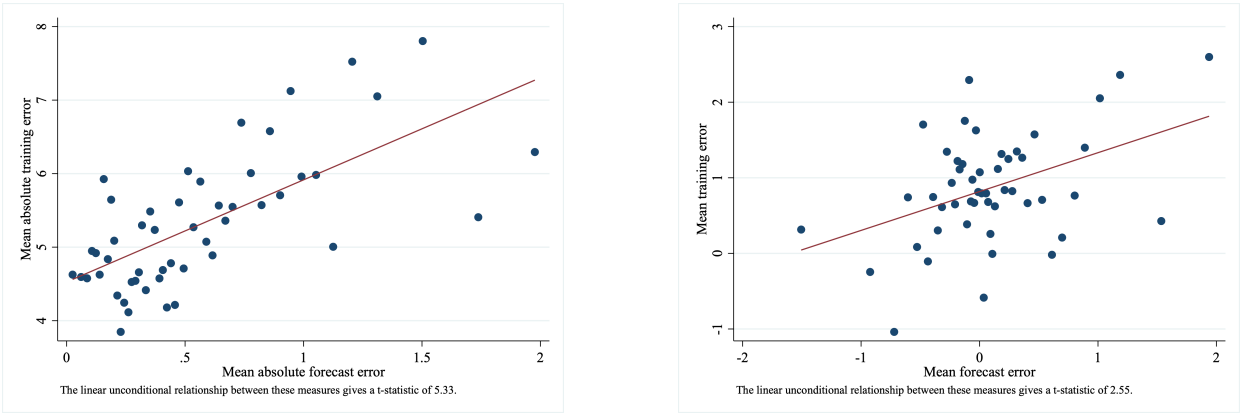


Figure A7. : Forecasting training accuracy is correlated with firms' prior forecast accuracy (left figure), and bias is correlated with firms' prior forecasting bias (right figure)

Note: Panel showing binscatters of the correlations between forecast training accuracy and bias and prior firms bias and accuracy. Data on training errors comes from round 7, 9, and 10 on 3,151 averaged firm responses, aggregating forecasts made for hypothetical businesses. The forecasting error (bias) is calculated as the average using 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months), while the absolute error is the absolute of that figure.

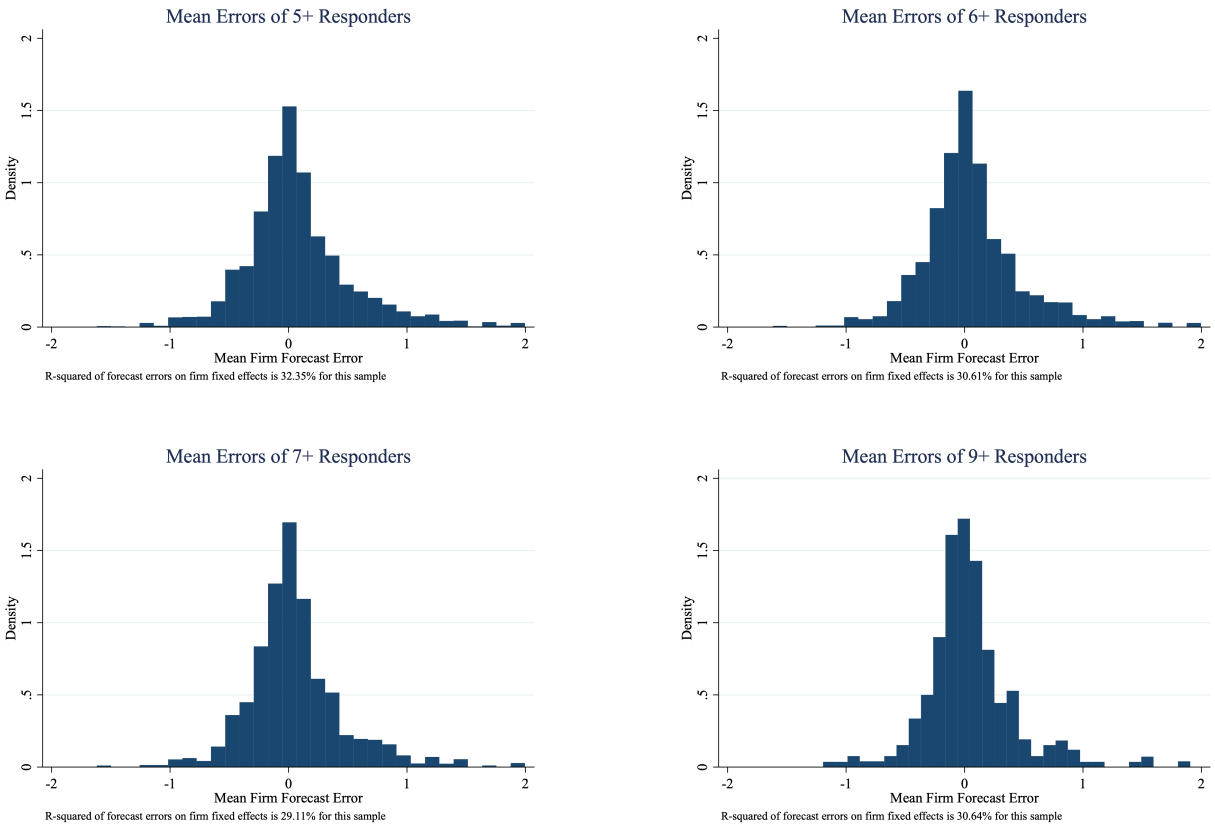


Figure A8. : Mean Forecast Errors for Different Response Thresholds

Note: Note: These plots report the average forecast error for firms, varying the minimum number of responses required for inclusion. The original figure in the main text uses a threshold of 8+ responses. Subfigure (a) includes firms with 5+ responses, (b) with 6+, (c) with 7+, and (d) with 9+. The mean firm level bias is calculated by averaging the forecasting error (arc-percentage) across all forecasts using the formula: 2*(Forecast - Actual)/(Forecast + Actual). All firms have equal weight.

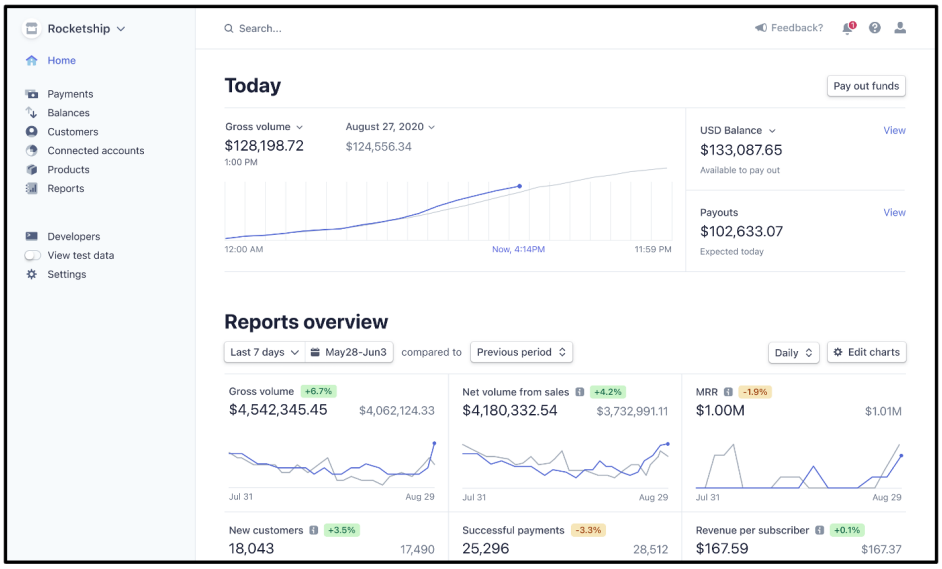


Figure A9. : A (fictional) example of the dashboard

Note: This figure shows a fictional example of the dashboard, which is similar to those we asked firms to access during the dashboard experiment. The experiment ran in waves 5 to 10, when firms were asked in the survey to use the dashboard to report their id and last quarter revenue.

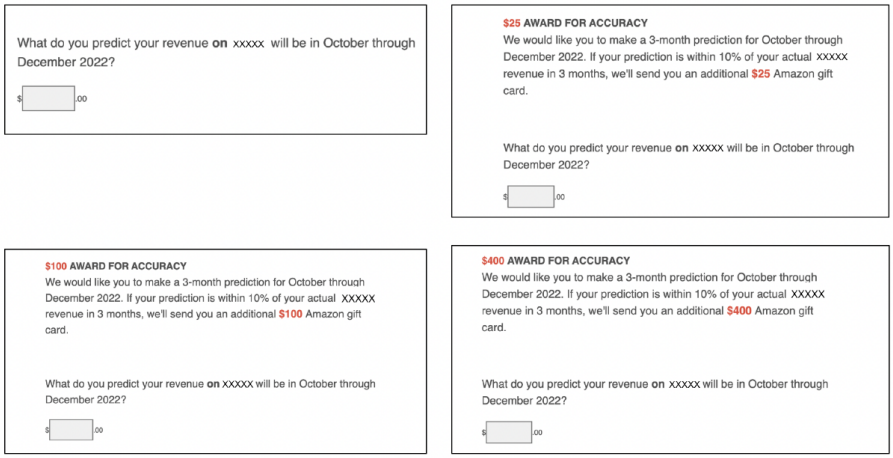
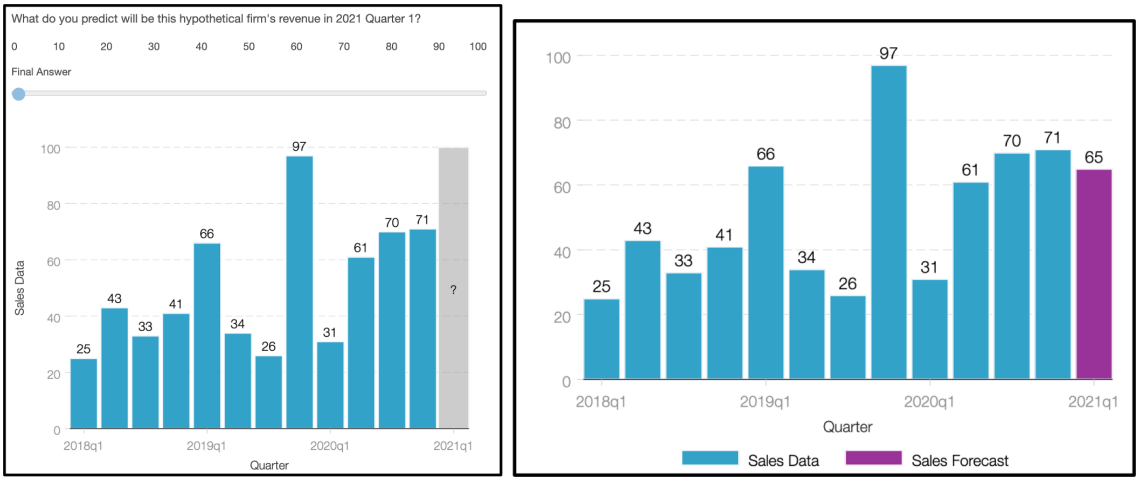


Figure A10. : Forecast payment randomization from $0 to $400, four examples

Note: This figure shows example prompts from the rewards randomization, which we ran in waves 5, 7 and 9 rewards were randomized from $0 (top left) to $400 (bottom right). In all other waves the reward was $25 (top right). The name of the Fintech has been replaced with XXXXX in these screenshots. Rewards were varied with incentives from $0-$400.

Figure A11. : Forecasting training exercise: An example



Note: On the left hand side, we show an example question from our module of hypothetical firm forecasts. On the right hand side, we show one of our firm sales forecasts for this firm. Our forecast was generated from a regression with industry dummies of current forecasts on last 4 quarters (in logs). This generated a predicted sales with weight of about 1/2 on last quarter, and weight of about 1/6 on prior three quarters.

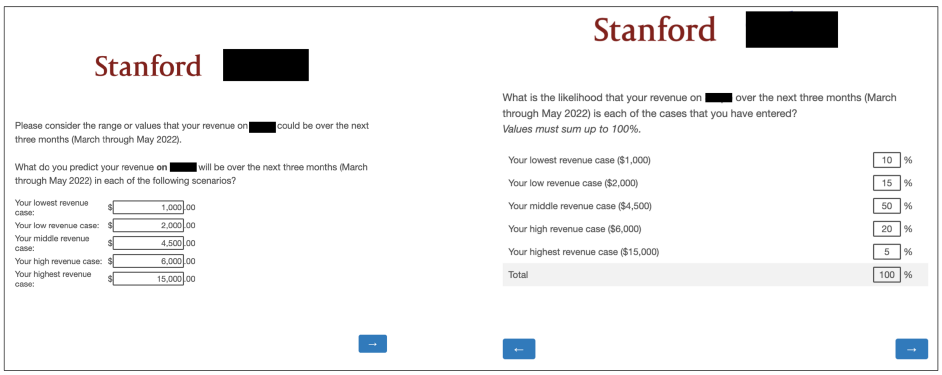


Figure A12. : Subjective sales forecast distribution question

Note: This figure shows a fictional example of the scenario questions, which is similar to those we asked firms to complete during the scenario experiment. The experiment ran in waves 8 and 10. The name of the Fintech has been replaced with a blurring box in these screenshots. Shown for a hypothetical firm.

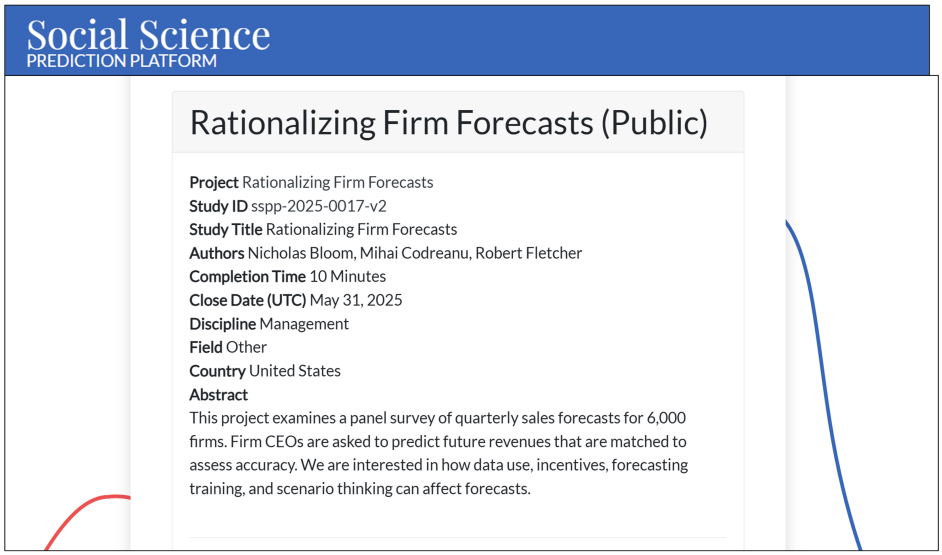


Figure A13. : We also collected economists' beliefs about our results

Note: This figure shows a screenshot from the welcome page of our study on the Social Science Prediction Platform, eliciting economists' beliefs about our study and experiments results. This survey was available at the following link: https://socialscienceprediction.org/predict/.

Table A1—: Predictors of Survey Response

Responding in a Subsequent Wave							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Responded This Wave						
$\text{Log}(1 + \text{Last Q Sales})_{i,t-1}$	0.012***	-0.012***	-0.011***	-0.009***	-0.006***	-0.010***	-0.023***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\text{Log}(1 + \text{Last Year Sales})_{i,t-1}$		0.027***	0.025***	0.028***	0.028***	0.029***	0.029***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\text{Log}(\text{Nb Employees})_{i,t-1}$				-0.030***	-0.034***	-0.031***	-0.020***
				(0.003)	(0.003)	(0.003)	(0.003)
$\text{Forecast Error}_{i,t-1}$						-0.014***	-0.013***
						(0.003)	(0.003)
$\text{Squared Forecast Error}_{i,t-1}$							-0.039***
							(0.003)
Industry FE	No	No	Yes	Yes	Yes	Yes	Yes
Wave by Round FE	No	No	No	No	Yes	Yes	Yes
R-squared	0.008	0.014	0.019	0.025	0.239	0.240	0.247
Observations	60692	60433	60433	58215	58215	58215	58215

Notes: This table presents predictors of survey participation as the probability of a firm that responded in a previous wave responding in the current wave. Column (1) includes lagged 3-month sales. Column (2) adds lagged 12-month sales. Column (3) adds industry fixed effects. Column (4) adds the number of employees. Column (5) adds wave-by-round fixed effects. Columns (6) and (7) add measures of the firm's past forecast errors: last forecast error and last forecast error squared. Standard errors are clustered at the firm level. Stars denote significance: * p<0.10, ** p<0.05, *** p<0.01.

Table A2—: Impact of Experiments on Future Response Rates

	(1) Valid Response Probability
Panel A: Dashboard	
$\text{Dashboard}_{i,t-1}$	0.046***
	(0.010)
N	8,786
Dependent Mean	0.496
Panel B: Rewards	
$\text{Rewards ('00s)}_{i,t-1}$	-0.004
	(0.005)
N	6,260
Dependent Mean	0.455
Panel C: Training	
$\text{Training}_{i,t-1}$	-0.006
	(0.015)
N	4,495
Dependent Mean	0.429
Panel D: Scenarios	
$\text{Scenarios}_{i,t-1}$	-0.011
	(0.027)
N	1,272
Dependent Mean	0.630

Notes: This table reports the dynamic impacts of the dashboard (Panel A), rewards (Panel B), training (Panel C) and scenarios (Panel D) treatments in the previous waves on valid response probabilities in the next wave. Dependent variables is an indicator taking value 1 if the firm has a valid response in the immediate next wave, and 0 otherwise. The sample is all firms that have responded at least once. We only include waves after experiment randomization. The dependent variable mean is reported as the control mean for the untreated individuals who responded in the previous wave. All regressions have full time fixed effects (consisting of the wave by round dummies), standard errors are clustered at the firm level. Stars * represent significance at 10% level i.e. p-value of mean being 0 is <0.1 ** p-value is <0.05 *** p-value is <0.01.

Table A3—: Forecasting Accuracy and Size

	(1) Winner	(2) Winner	(3) Winner	(4) MSE	(5) MSE	(6) MSE	(7) Forecast Error	(8) Forecast Error	(9) Forecast Error
Log (Nb Employees)	0.016***	0.008**	0.010***	-0.064***	-0.043***	-0.049***	0.007	0.011	0.011
	(0.003)	(0.003)	(0.003)	(0.013)	(0.013)	(0.013)	(0.008)	(0.008)	(0.009)
Wave by Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
FinTech Share Controls	No	No	Yes	No	No	Yes	No	No	Yes
R-squared	0.014	0.024	0.032	0.045	0.056	0.065	0.024	0.030	0.030
Observations	17556	17556	17556	17556	17556	17556	17556	17556	17556

Notes: This table reports the relationship between winning percentages, accuracy (mean squared error), forecast errors (bias) and firm size (log of employment). In Columns (1)-(3), the dependent variable is the winner variable, defined as 1 if the firm was a winner in the particular round and 0 otherwise. In Columns (4)-(6), the dependent variable is the squared of the forecast error. In Columns (7)-(9) dependent variable is the forecast error, defined as 2*(Forecast - Actual)/(Forecast + Actual). In columns (1), (4) and (7) we use use as controls only round by wave fixed effects. In columns (2), (5) and (8), we also introduce as controls industry fixed effects. In columns (3), (6) and (9), we also use as controls percentage of revenues on FinTech. Stars denote significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A4—: Forecast Errors Predict Employment

	(1) Log Employees	(2) Log Employees	(3) Log Employees	(4) Log Employees	(5) Log Employees	(6) Log Employees
Log(1 + Last Q Sales)$_{i,t}$	0.233***	0.242***	0.037***	0.122***	0.152***	0.034***
	(0.007)	(0.007)	(0.005)	(0.011)	(0.012)	(0.006)
Log(1+Last Q Sales)$_{i,t-1}$				0.170***	0.149***	0.022***
				(0.012)	(0.012)	(0.006)
Forecast Error$_{i,t}$		0.132***	0.030***		0.102***	0.027***
		(0.012)	(0.005)		(0.015)	(0.007)
Forecast Error$_{i,t-1}$					0.133***	0.014*
					(0.017)	(0.008)
Period by Wave FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes		Yes	Yes	
Firm FE			Yes			Yes
R-squared	0.295	0.304	0.908	0.356	0.372	0.928
Observations	17,556	17,556	14,837	9,039	9,039	7,988

Notes: This table reports the relationship between forecast errors and last quarter sales and employment. The dependent variable is the natural logarithm of the total number of employees of the firm (including full-time and part-time). The independent variables are the size of the firm in the 3-months previous to the prediction (the natural logarithm of the sales in last 3 months), and its lag from the previous response, the forecast error from the current quarter 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months+Actual Sales Next 3 Months), as well as its lag measured using the arc-percentage definition. All regressions include industry and period by wave fixed effects, while columns (3), (6) also include firm fixed effects. Stars represent significance at 10% level i.e. p-value of mean being 0 is <0.1 p-value is <0.05 p-value is <0.01.

Table A5—: Robustness Checks for Dashboard Treatment

Specification	(1) Bias	(2) MSE	(3) Predictability	(4) Precision	(5) Win Rate
Baseline	-0.037**	-0.097***	-0.050	-0.033***	0.020**
	(0.016)	(0.022)	(0.038)	(0.010)	(0.008)
R-squared	0.013	0.023	0.067	0.011	0.006
N	9,994	9,994	5,594	9,810	9,994
Dep. Mean	0.110	0.676	0.089	0.410	0.172
High FinTech Sample	-0.045**	-0.085***	-0.041	-0.034***	0.020**
	(0.019)	(0.026)	(0.048)	(0.012)	(0.010)
R-squared	0.016	0.017	0.081	0.012	0.011
N	6,549	6,549	3,662	6,438	6,549
Dep. Mean	0.128	0.602	0.101	0.410	0.191
Bigger Firms Sample	0.003	-0.031	-0.080	-0.050***	0.038**
	(0.020)	(0.025)	(0.072)	(0.017)	(0.017)
R-squared	0.019	0.013	0.119	0.021	0.025
N	2,873	2,873	1,742	2,846	2,873
Dep. Mean	0.001	0.303	0.028	0.408	0.254
Firm Fixed Effects	-0.022	-0.072***	-0.053	-0.026**	0.011
	(0.018)	(0.023)	(0.042)	(0.012)	(0.010)
R-squared	0.459	0.542	0.483	0.365	0.393
N	7,999	7,999	4,784	7,853	7,999
Dep. Mean	0.080	0.619	0.068	0.401	0.194
Attrition Weights	-0.043**	-0.122***	-0.059	-0.038***	0.014
	(0.019)	(0.027)	(0.040)	(0.011)	(0.009)
R-squared	0.019	0.038	0.074	0.013	0.005
N	8,367	8,367	5,594	8,232	8,367
Dep. Mean	0.108	0.677	0.089	0.411	0.182

Notes: This table reports robustness checks for the Dashboard Treatment. Each column corresponds to a separate regression of the outcome variable listed at the top. The rows correspond to different model specifications. For each specification, we report the coefficient on the treatment variable, with the standard error in parentheses below it. Subsequent rows show the adjusted R-squared, number of observations, and the mean of the dependent variable for the estimation sample. Specifications are: (1) Baseline: includes wave-by-round fixed effects. (2) High FinTech Sample: restricts the baseline to firms having at least 50% or more revenue from FinTech. (3) Bigger Firms: includes only firms earning annualized sales of above $100,000/year. (4) Firm Fixed Effects: includes firm fixed effects. (5) Attrition Weights: re-weights the baseline specification using inverse probability weights. All standard errors are clustered at the firm level. *, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.

Table A6—: Robustness Checks for Rewards Treatment

Specification	(1) Bias	(2) MSE	(3) Predictability	(4) Precision	(5) Win Rate
Baseline	-0.019***	-0.009	0.010	0.005	0.009**
	(0.007)	(0.010)	(0.019)	(0.005)	(0.004)
R-squared	0.005	0.005	0.069	0.006	0.005
N	6,260	6,260	2,668	6,150	6,260
Dep. Mean	0.089	0.609	0.056	0.374	0.164
High FinTech Sample	-0.027***	-0.009	0.002	0.003	0.008*
	(0.009)	(0.012)	(0.026)	(0.006)	(0.005)
R-squared	0.008	0.006	0.089	0.012	0.010
N	4,090	4,090	1,710	4,022	4,090
Dep. Mean	0.110	0.555	0.065	0.377	0.180
Bigger Firms Sample	-0.009	-0.002	-0.014	0.011	0.016**
	(0.009)	(0.011)	(0.035)	(0.009)	(0.008)
R-squared	0.012	0.010	0.126	0.016	0.018
N	1,772	1,772	856	1,757	1,772
Dep. Mean	-0.021	0.305	-0.007	0.357	0.240
Firm Fixed Effects	-0.016	-0.024*	-0.032	-0.008	0.024***
	(0.011)	(0.013)	(0.025)	(0.008)	(0.006)
R-squared	0.542	0.608	0.580	0.506	0.507
N	3,748	3,748	1,640	3,688	3,748
Dep. Mean	0.031	0.499	0.012	0.365	0.194
Attrition Weights	-0.019**	-0.011	0.009	0.003	0.010**
	(0.009)	(0.012)	(0.020)	(0.006)	(0.004)
R-squared	0.007	0.010	0.072	0.007	0.004
N	4,686	4,686	2,668	4,622	4,686
Dep. Mean	0.079	0.601	0.056	0.374	0.172

Notes: This table reports robustness checks for the Rewards Treatment. Each column corresponds to a separate regression of the outcome variable listed at the top. The rows correspond to different model specifications. For each specification, we report the coefficient on the treatment variable, with the standard error in parentheses below it. Subsequent rows show the adjusted R-squared, number of observations, and the mean of the dependent variable for the estimation sample. Specifications are: (1) Baseline: includes wave-by-round fixed effects. (2) High FinTech Sample: restricts the baseline to firms having at least 50% or more revenue from FinTech. (3) Bigger Firms: includes only firms earning annualized sales of above $100,000/year. (4) Firm Fixed Effects: includes firm fixed effects. (5) Attrition Weights: re-weights the baseline specification using inverse probability weights. All standard errors are clustered at the firm level. *, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.

Table A7—: Robustness Checks for Training Treatment

Specification	(1) Bias	(2) MSE	(3) Predictability	(4) Precision	(5) Win Rate
Baseline	-0.009	0.060*	0.072	0.013	-0.014
	(0.023)	(0.032)	(0.064)	(0.014)	(0.011)
R-squared	0.018	0.031	0.083	0.010	0.006
N	5,703	5,703	2,494	5,586	5,703
Dep. Mean	0.096	0.630	0.056	0.387	0.180
High FinTech Sample	-0.003	0.090**	0.087	0.015	-0.014
	(0.027)	(0.038)	(0.079)	(0.017)	(0.014)
R-squared	0.018	0.020	0.107	0.012	0.012
N	3,751	3,751	1,633	3,679	3,751
Dep. Mean	0.109	0.544	0.063	0.389	0.200
Bigger Firms Sample	0.001	0.039	0.155	0.016	-0.004
	(0.029)	(0.036)	(0.117)	(0.026)	(0.023)
R-squared	0.018	0.013	0.169	0.017	0.027
N	1,603	1,603	820	1,588	1,603
Dep. Mean	-0.028	0.268	0.001	0.381	0.265
Firm Fixed Effects	0.031	0.035	0.127	0.047**	-0.037**
	(0.031)	(0.038)	(0.086)	(0.020)	(0.017)
R-squared	0.538	0.620	0.547	0.476	0.493
N	3,700	3,700	1,752	3,625	3,700
Dep. Mean	0.051	0.559	0.039	0.379	0.211
Attrition Weights	0.007	0.073*	0.082	0.024	-0.011
	(0.029)	(0.043)	(0.067)	(0.016)	(0.012)
R-squared	0.029	0.053	0.094	0.014	0.004
N	4,308	4,308	2,494	4,231	4,308
Dep. Mean	0.092	0.634	0.056	0.387	0.192

Notes: This table reports robustness checks for the Training Treatment. Each column corresponds to a separate regression of the outcome variable listed at the top. The rows correspond to different model specifications. For each specification, we report the coefficient on the treatment variable, with the standard error in parentheses below it. Subsequent rows show the adjusted R-squared, number of observations, and the mean of the dependent variable for the estimation sample. Specifications are: (1) Baseline: includes wave-by-round fixed effects. (2) High FinTech Sample: restricts the baseline to firms having at least 50% or more revenue from FinTech. (3) Bigger Firms: includes only firms earning annualized sales of above $100,000/year. (4) Firm Fixed Effects: includes firm fixed effects. (5) Attrition Weights: re-weights the baseline specification using inverse probability weights. All standard errors are clustered at the firm level. *, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.

Table A8—: Robustness Checks for Scenarios Treatment

Specification	(1) Bias	(2) MSE	(3) Predictability	(4) Precision	(5) Win Rate
Baseline	0.022	0.005	0.007	0.023	-0.006
	(0.034)	(0.050)	(0.066)	(0.020)	(0.016)
R-squared	0.018	0.039	0.061	0.003	0.007
N	2,480	2,480	1,931	2,422	2,480
Dep. Mean	0.175	0.759	0.120	0.435	0.198
High FinTech Sample	-0.031	-0.057	0.093	0.032	0.004
	(0.037)	(0.057)	(0.081)	(0.025)	(0.021)
R-squared	0.019	0.028	0.064	0.004	0.009
N	1,657	1,657	1,303	1,621	1,657
Dep. Mean	0.211	0.671	0.159	0.419	0.223
Bigger Firms Sample	0.057	0.018	0.187	0.040	0.037
	(0.041)	(0.056)	(0.152)	(0.038)	(0.034)
R-squared	0.011	0.015	0.076	0.019	0.037
N	703	703	564	694	703
Dep. Mean	0.039	0.271	0.041	0.427	0.279
Firm Fixed Effects	0.072	0.064	-0.020	0.001	0.017
	(0.048)	(0.064)	(0.097)	(0.036)	(0.028)
R-squared	0.652	0.683	0.648	0.544	0.593
N	1,352	1,352	990	1,310	1,352
Dep. Mean	0.105	0.569	0.087	0.415	0.232
Attrition Weights	0.032	0.016	-0.048	0.023	-0.011
	(0.039)	(0.059)	(0.070)	(0.021)	(0.016)
R-squared	0.024	0.054	0.074	0.004	0.006
N	2,444	2,444	1,931	2,388	2,444
Dep. Mean	0.163	0.735	0.120	0.433	0.201

Notes: This table reports robustness checks for the Scenarios Treatment. Each column corresponds to a separate regression of the outcome variable listed at the top. The rows correspond to different model specifications. For each specification, we report the coefficient on the treatment variable, with the standard error in parentheses below it. Subsequent rows show the adjusted R-squared, number of observations, and the mean of the dependent variable for the estimation sample. Specifications are: (1) Baseline: includes wave-by-round fixed effects. (2) High FinTech Sample: restricts the baseline to firms having at least 50% or more revenue from FinTech. (3) Bigger Firms: includes only firms earning annualized sales of above $100,000/year. (4) Firm Fixed Effects: includes firm fixed effects. (5) Attrition Weights: re-weights the baseline specification using inverse probability weights. All standard errors are clustered at the firm level. *, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.

Table A9—: Effects of Treatments on Firm Performance

	(1) Labor Productivity	(2) Labor Productivity	(3) Growth Next Year	(4) Growth Next Year	(5) Log Sales (next yr)	(6) Log Sales (next yr)	(7) Survival (7-12 mo.)	(8) Survival (7-12 mo.)
Squared Forecast Error$_{i,t}$	-0.726***	-0.316	-0.178***	-0.192	-1.031***	-0.261	-0.067***	-0.026
	(0.032)	(0.506)	(0.010)	(0.153)	(0.043)	(0.620)	(0.004)	(0.065)
Instrument		Dash+Rew		Dash+Rew		Dash+Rew		Dash+Rew
First stage F-stat		6.577		6.853		6.514		6.514
R-squared	0.155	0.097	0.106	0.090	0.201	0.083	0.085	0.046
N	8,410	8,410	8,355	8,355	8,786	8,786	8,786	8,786
Dep. Mean	9.529	9.529	-0.083	-0.083	10.449	10.449	0.937	0.937

Notes: This table reports the regression results from the squared forecast error on revenue labor productivity and growth (columns 1-4) and sales and survival (columns 5-8). The independent variable in all the subpanels is [2*(Predicted Sales Next 3 Months-Actual Sales Next 3 Months)/(Predicted Sales Next 3 Months+Actual Sales Next 3 Months)]ˆ2. In Panel A, the dependent variables are: revenue labor productivity measured as the log(Sales Next 12 Months)/Number of Employees (columns 1-2); growth in the next year measured as [2*(Sales Next 12 Months- Sales Last 3 Months)/(Sales Next 12 Months+Sales Last 3 Months)] (columns 3-4). In Panel B, the dependent variables are: log size, measured as the log of the sales in the next year (column 5-6); and survival (having any income 7-12 months after the survey). Instrumental variables in even columns are the dashboard and reward treatments. All regressions are run for round 5-9, due to absence of 12-month data in round 10. All regressions also have reported the R-squared, the number of observations (when the dependent variables are defined). The dependent variable mean is reported as the control mean for the untreated individuals. All regressions have full time fixed effects (consisting of the period and wave dummies), standard errors are clustered at the firm level. Stars represent significance at 10% level i.e. p-value of mean being 0 is <0.1 p-value is <0.05 p-value is <0.01.

Table A10—: Model Parameters

Panel A: Common Parameters		
Parameter	Symbol	Value
Capital Elasticity	α	0.30
Quadratic Adjustment Cost	ξ	0.5
Interest Rate (Quarterly)	r	0.0259
Depreciation Rate (Quarterly)	δ	0.0259
True Productivity Persistence	ρ	0.9
Productivity Shock Std. Dev.	σ_{ε}	0.1
Mean Productivity	$\bar{A}$	1.0
Output Price	p	0.1
Number of Firms	N	5,000
Number of Periods	T	1,000
Panel B: Behavioral Parameters		
	Expected Biases	**Actual Biases**
Mean Optimism Bias	0.096	0.163
Std. Dev. Optimism Bias	0.08	0.08
Perceived Persistence ($\rho^{perceived}$)	0.92	0.94
$P(l = 0)$ Current Info	0.75	0.50
$P(l = 1)$	0.10	0.20
$P(l = 2)$	0.05	0.10
$P(l = 3)$	0.05	0.10
$P(l = 4)$	0.05	0.10

Notes: This table reports the parameters used in the firm investment model described in Section 5. Panel A reports parameters common to all model specifications. Each firm produces output $Y_{it} = pA_{it}K_{it}^{\alpha}$ using capital K_{it} and faces stochastic productivity shocks A_{it} following an AR(1) process: $A_{i,t+1} = (1-\rho)\bar{A} + \rho A_{it} + \varepsilon_{it+1}$, where $\varepsilon_{it+1} \sim N(0, \sigma_{\varepsilon}^2)$. The firm solves: $V(K, A) = \max_{K'}\{pAK^{\alpha} - (r+\delta)K - C(K', K) + \frac{1}{1+r}E^{F}[V(K', A')]\}$ where $C(K', K) = \frac{\xi}{2}\frac{(K'-(1-\delta)K)^2}{K}$ is the quadratic adjustment cost on net investment. The interest rate $r = 0.0259$ corresponds to approximately 10% annually. Panel B reports behavioral parameters calibrated to match the forecasting biases documented in Section III. "Expected Biases" are based on the Social Science Prediction Platform survey results. "Actual Biases" are based on the average biases found in the firm-level data. Over-optimism is modeled as a firm-specific, time-invariant additive bias $b_i \sim N(\mu_b, \sigma_b^2)$ on perceived productivity. Predictable errors represent inattention, where $P(l = k)$ is the probability that a firm uses productivity information from k periods ago. Over-precision is modeled through perceived persistence $\rho^{perceived} > \rho$, reflecting firms' belief that productivity shocks are more persistent than they actually are.

Data and Experimental Details

B1. Data cleaning

In this Appendix, we provide more details over the survey collection and analysis. First, we create a sample frame according to the details provided in Section II.A, of active for-profits US businesses on FinTech platform with a sufficient history and volume of revenue. They were invited for a first baseline survey (collecting more information about the business, CEO characteristics, industry, etc. as well as the forecasting exercises) or a follow-up survey (that is more streamlined across the experiments and forecasting questions). These survey waves also had rounds. Each wave, a small pilot round of about 50 firms was sent out first as a test sample. This checked for survey errors, bugs or other issues. The following months, 1–5 other rounds (containing the bulk of the firms) were sent in waves a couple of days apart each for capacity constraints (capacity issues came from the fact requests generated some replies that needed to be manually fixed – like questions on payments, confirmations this was a real survey etc – and to avoid getting caught by spam filters for sending many 1000s of emails). All time fixed effects (unless otherwise noted) account by full wave by round dummies. The exact timing of the waves was determined by logistical and data availability reasons. The plan was for them to be exactly 6 months apart (to leave 3 months for the survey period and 3 months for follow-up) but some timing was shifted by a few months back or forwards. For example, one wave was delayed when FinTech's Stata permissions expired (which was required for the sample creation). Most rounds also had 2–3 reminders in addition to the initial invite, that we sent at different times of the week and hours of the day to maximize response rates.

For the data to appear in our final sample we drop firms that did not provide a prediction, that had no transactions in the past 3 months, that rushed through the survey (defined as finishing it in less than 7 minutes, which was about 70 percent the minimum time the coauthors and FinTech employees knowing the survey needed to complete it), those that did not have, on average, at least 5 percent of their revenue on FinTech and those that were closed or expecting their operations to be closed within 3 months. To put numbers onto these restrictions, throughout all waves, our initial sample had 25,164 responses from firms that finished or did not finish the survey but provided a prediction. Dropping those that did not have any sales in the last 3 months prior to the survey got the sample to 22,430 observations. Then, we dropped 2,231 more observations that rushed through the survey, leaving 20,199 valid responses. Finally, our sample got to 18,363 after excluding firms that had a very low percentage of FinTech (< 5 percent) and then excluded 303 additional firms that were closed and predicting they were closed at the time of taking the survey (had 0 actuals and predictions). The final sample we worked with was 18,060 responses. We also had a small number of firms that had missing descriptive variables as shown in Table 1; for the CEO sample,

where possible, we input CEO characteristics based on answers in any waves. For 3 firms that did not answer the percentage of their sales happening online, we used their FinTech percentages as a replacement. For the 13 firms that had missing total sales as reported, we input them using the actual sales that happened on FinTech and scaling them upwards by the percentage declared they have on FinTech. In the main body of this paper, we also winsorized the time spent taking the survey and the time spent on the forecasting question. The raw total duration is calculated as the time the survey is submitted minus the time the survey started. Due to a sizable number of respondents who stopped while taking the survey creating outliers (many responses have durations of days or weeks, presumably due to starting completing the survey and then stopping and coming back when reminders are sent) in the raw data, we decided to winsorize this variable at the 30 minutes mark, which is about twice what we expected a survey to take.[39] For the time spent on the forecasting question, this records the time spent on the forecasting page (which may be slightly different between waves). Due to a similar logic (presence of outliers presumably due to stopping), we winsorized this at the 10 minutes mark, which is about the 99th percentile of the forecasting time. All results on timing outcomes are generally robust to being either more restrictive or lax on winsorization.

B2. Experiments

The four experiments were cross-randomized, before sending the survey invites, according to the last digits of the account numbers. These were considered as good as random. The four interventions in this paper are described as following:

Dashboard: before making their prediction in the main forecasting exercise, treated firms are requested to log in their account, copy and paste the last 6 digits of your account ID, open the dashboard, and select the relevant last 3 months to provide their FinTech revenues. A (fictional) example of a dashboard is shown in Appendix Figure A9. Firms were not compelled/stopped from continuing the survey if they give the wrong 6 digits, Therefore, the interpretation of the results of this intervention can be interpreted as an intent-to-treat.

Rewards: on the page making their prediction in the main forecasting exercise, firms are shown varying rewards amounts between \$0 (no reward) to \$400: 25 percent of respondents were randomly assigned \$0, another 25 percent continued with the base amount of \$25, and the remaining 50 percent were randomly placed into eight equally sized groups of \$50, \$100, \$150, \$200, \$250, \$300, \$350, and \$400. Examples of the questions (with the varying rewards

[39] Even winsorizing at the 90th percentile would not take care of this issue, as some waves have 20-25% of responses clearly stopping between first starting and completing the survey (close to an hour).

amounts) that were shown to firms are in Appendix Figure A10. These rewards, as well as participation payments) were sent out (in the form of Amazon gift vouchers) before the beginning of the next survey round to maximize salience.

Training: before making their prediction in the main forecasting exercise, treated firms are requested to analyze data from up to 10 vignettes of hypothetical firms (examples in Figure A11), and are then shown our own forecast. We randomized the sample into three groups, each of about a third: one third saw the training exercises before making the prediction, one third after (to maintain a roughly constant total survey length and thinking this treatment may have some dynamic effects), and one third did not see it at all. The results in the main text are shown for the group that saw the training before making the prediction compared to both groups that did not. Note that we obtain similar (null) treatments results comparing the before vs after group and omitting the group that did not see the survey at all. For the reason of these null effects, we simplified the exposition in the maint text.

Scenarios: before making their prediction, treated firms are requested to provide 5-point distribution of potential sales across the lowest, low, middle, high and highest sales scenarios and their associated probabilities (example in Appendix Figure A12). For the purposes of creating the variance of this distribution, the coefficient of variation, and measures of over-precision, we also elicited distributions for control firms, as well as simplified 3-point distributions (worst, middle, best) for waves 5, 6, 7, and 9, along with their associated probabilities after they made their rewarded forecasts. These variables were used alongside for creating subjective distributions and comparing their variance to the actual recorded distributions for the whole sample.

B3. Subjective Distributions

To translate firms' three- and five-point scenario forecasts into a pair of comparable empirical distributions—one for beliefs and one for outcomes—we adopt the workflow similar to papers such as Altig et al. (2020). We begin by screening the data so that every respondent contributes a "proper" probability distribution. A response is dropped whenever (i) an individual bin receives the entire 100 percent probability mass, (ii) the stated probabilities do not sum to 100, or (iii) all bin endpoints coincide, indicating a degenerate forecast. We also ensure that respondents are required by the survey software to put increasing values of sales in bins corresponding to "better" scenarios. For every admissible response we first calculate its probability-weighted mean forecast. We then express each scenario's sales level as an arc-percentage change relative to that mean. This transformation places gains and losses on a common footing, centers the distribution at zero for every firms and confines all values to lie between minus two and plus two—properties that prevent

extreme levels from dominating subsequent dispersion statistics. This is also the same type of transformation used for the forecast error throughout the paper.

With the scenarios on a common scale, we harmonize labels across waves—five-bin questionnaires keep "lowest" through "highest," whereas three-bin waves are relabeled "low," "mid," and "high"—and carry the associated probabilities along. We only asked these questions during waves 5–10, with waves 8 and 10 asking the 5-bins questions, and the rest the 3-bin questions. We next reshape the data so that every probability-growth pair becomes its own record and attach a weight equal to the reported probability share. Stacking this weighted set across all firms and waves yields an empirical density that is, by construction, the aggregate subjective distribution of expected three-month-ahead sales growth, relative to the mean.

For each firm-wave we compute the actual three-month-ahead growth rate on the same arc scale and append it as a single, equally weighted observation for all firms. Overlaying histograms of the weighted beliefs and the unweighted outcomes gives an immediate visual read on the centre, spread, and tails of subjective expectations versus realized performance. This is what is shown in Figure 4.

Throughout the paper, we use two additional (simple) ways to summarize each respondent's uncertainty. First, we form a variance on the arc scale by summing, across bins, the probability shares times the squared arc deviation from the mean, reported in Table 2. While very similar to the counterpart in Figure 4, this has the advantage of being calculated without the approximations and stacking procedures described above (i.e., the need to assign a representative point to each interval and pool observations across firm x wave x bins). The results are, nonetheless, very similar. Second, we follow previous papers and repeat the variance calculation in original currency units and divide its square root by last-quarter revenue, producing the firm's coefficient of variation. This latter measure is winsorized at the fifth and ninety-fifth percentiles to temper the influence of outliers and is used in Figure 9.

B4. Management Practices

Some firms were asked to respond to a module on management practices. The questions were copied with minor adjustments from the list of questions in Bloom et al. (2019), used by the Management and Organizational Practices Survey (MOPS) and the Annual Survey of Entrepreneurs (ASE) surveys. These are some examples of the questions asked, where answers were scored and (where appropriate) normalized around the mean according to the algorithm used by the Census:

1) How many key performance indicators (KPIs) are monitored at your business?

2) How frequently are KPIs typically reviewed at your business?

3) What do you do when a service or production problem arises in your business?

4) What describes the time frame of your service/production targets?

5) How easy or difficult is it to achieve service, or production targets?

6) What are the primary ways employees are promoted in your business?

7) When is an under-performing employee reassigned or dismissed?

AEA RCT Registry

Our AEA RCT Registry was published on January 21, 2021, with the ID AEARCTR-0007058. All the information following is reproduced verbatim from the initial submission.

Abstract: In a panel survey of online firms, we find that the ability to correctly forecast sales is highly correlated with firm performance. Despite its apparent importance, firms are remarkably bad at forming accurate predictions – if firms simply accurately reported their sales over the past 3-months as their prediction they would perform twice as well. We posit that simply encouraging them to review the financial data already easily available to them would dramatically improve their prediction performance. We aim to run an RCT testing exactly that and show a causal pathway from monitoring business financials to prediction performance.

Keywords: Firms & Productivity

Additional Keywords: management, uncertainty, sales

Intervention(s): In conjunction with our partner, FinTech, we have been tracking firms and their forecast abilities over time by comparing 3-month predictions for revenue on FinTech with their actual sales on FinTech. In order to ensure that firms are offering serious predictions, we reward them with a $25 gift card if their predictions are within 10% of their actual sales. We are testing two interventions for significant effects on their predictions: having firms review their prior historical data on the FinTech dashboard before making their prediction and varying the amount of the gift card they can earn with their predictions. The first intervention tests whether or not having firms review their financial details on the FinTech dashboard as part of the survey prior to forming forecasts can significantly increase their forecast accuracy. In particular, we have them sign-in to their FinTech dashboard and report their revenue over the last 3-months before making their predictions. On top of this, we are also testing whether or not changing firms earns with a correct prediction can incentivize them to provide more accurate responses. We vary the amounts from $0 to $400.

Primary Outcomes (end points): Forecast accuracy, winning prediction (i.e. within 10%), prediction bias, time spent forming prediction, checking their financial dashboard

Secondary Outcomes (end points): Productivity, growth rates, revenue, survival (having transaction in a 6 month period).

Experimental Design: Our interventions are cross randomized, with the dashboard treatment being run twice. For the first dashboard intervention, we randomly assigned 50% of the firms to control and 50% of the firms to treatment. While taking our quarterly survey, the treatment firms are asked to sign-in to their accounts and report their revenue over the last 3 months in the survey form. Control firms are simply asked to report their revenue and are not asked to consult their financial history. Both treatment and control then continue on to the question on 3-month predictions at that point. On the next round of the survey, 3–4 months later, we cross randomize again so that 25% are dashboard treatment twice, 25% are treatment then control, 25% are control then treatment, and 25% are control twice. Once at the prediction question, firms are randomly offered different amounts of money if their prediction for the next 3 months is within 10% of their actual revenue. Approximately 25% of the firms aren't offered a prize and 25% are offered $25 to match what we have previously offered them. The remaining 50% are evenly split into 6.25% shares for rewards of 50, 100, 150, 200, 250, 300, 350, and 400.

Randomization Method: We used the last few characters of their accounts ID numbers to assign sort them into their treatment arm. These account ID numbers are as good as random.

Randomization Unit: Firm

Was the treatment clustered? No

Sample size planned number of clusters: 8549 Firms

Sample size planned number of observations: 8549 Firms

Sample size (or number of clusters) by treatment arms: 1st Dashboard Treatment: 4271 Treatment, 4278 Control 2nd Dashboard Treatment: 4274 Treatment, 4275 Control Reward Treatment: $0 - 1860, $25 - $2289, $50 - 504, $100 - 618, $150 - 459, $200 - 637, $250 - 501, $300 - 510, $350 - 521, $400 – 650.

The two other experiments in this paper – the training and contingent thinking interventions – were added later, after the submission of the AEA RCT registry. So, these experiments were not pre-registered, but they were evaluated on the same metrics (preregistered outcomes) as the two pre-registered data and rewards experiments.

Social Studies Prediction Platform (SSPP) Data Collection

The results presented in this paper surprised us. So, we also ran our study on the Social Studies Prediction Platform, to elicit economists' priors about our study, including their beliefs about the effectiveness of the interventions. Our study collected responses between April 9, 2025 and May 31, 2025, and was available at the following link: https://socialscienceprediction.org/predict/. To encourage responses, we offered a lottery based $100 flat rate of participation incentive which would be gifted to the winners' charity of choice (No Kid Hungry was the winning charity). As we ran this data collection after the public release of the NBER Working Paper, we did not want to offer direct monetary incentives. We also posted the link to the SSPP platform and our study online, on LinkedIn, for a wider audience to sign-up and complete it. The final sample included 36 participants coming organically from the platform and 23 from LinkedIn that were quite similar in almost all substantive question responses. We had to exclude 6 responses from individuals who said they have read our paper before or seen it presented.

All the information following is reproduced verbatim from the initial submission, focusing on the key questions which was what a participant would have seen on the pages on the SSPP platform, when they completed the survey. A screenshot of the welcome page for the SSPP participants taking our survey is attached in Appendix Figure A13. To make sure not to induce any bias, we tried randomizing numbers, both positive and negative when we gave examples. These will be presented below in bolded and with the clear "random" mark. Because in small enough numbers the arc-percentage errors and the percentage points errors are quite similar, we simplified the questions and just asked about percentage points, so to not confuse any respondents. Note that we did not report here two sets of questions that were about the effects of treatment on optimism and over-precision (being outside of the lowest, highest bounds). The reason for this exclusion is that a number of respondents (about half) provided response predictions indicating an increase in the bias from the treatments. We believe from additional testing this was because of the double negative in the question confused respondees – the fact we asked about the impact of the treatment on the optimism (over-precision) required a negative number response to predict a reduction in the optimism (over-precision). As a result, the mean predicted impact was close to (and not significantly different from) 0, for all experiments, including the dashboard and rewards experiments which did improve some of these outcomes.

Abstract: This project examines a panel survey of quarterly sales forecasts for 6,000 firms. Firm CEOs are asked to predict future revenues that are matched to assess accuracy. We are interested in how data use, incentives, forecast training, and scenario thinking can affect forecasts. We will be donating $100 to a charity on behalf of a

randomly selected forecaster.

Authors: Nicholas Bloom, Mihai Codreanu, Robert Fletcher

Discipline: Management

Completion Time: 10 Minutes

Close Date (UTC): May 31, 2025

Background: This project examines a panel survey of quarterly sales forecasts for 6,000 firms. Firm CEOs are asked to predict future revenues that are matched to assess accuracy. The average firm has $750,000 of revenue per year and 6 employees.

An experiment with four interventions was conducted: Data Use - Firms are prompted to check a financial dashboard containing their recent revenues prior to making forecasts. Incentives - Firms were offered up to $400 for accurate forecasts. Forecasting Training - Firms were trained in making forecasts for ten hypothetical businesses similar to the survey sample. Scenario Thinking - Firms were asked about their lowest, low, middle, high, and highest outcomes BEFORE the revenue forecasting, along with their likelihood.

The control group are firms that are not assigned any of the above treatments. Firms were offered $50 for a survey, and an additional $25 if their forecast was within 10% of the actual value. Additionally, control firms were asked about their lowest, low, middle, and high outcomes AFTER the revenue forecasting, along with their likelihood. We are interested in your predictions of the treatment effects of each of the interventions.

Baseline Questions (each different pages):

A winner is defined as a firm that has completed the survey and whose forecasts were within 10% of the actual value. For example, if a firm had revenues of $100,000, then being within 10% would include forecasts between $90,000 and $110,000. So, a forecast of $95,000 would win. Among firms in the control group, please predict the win rate as a percentage. For example, a value of "random" means you are forecasting that "random"% of the firms in the control were able to forecast future revenues within 10% of their actual realized value.

The average forecasting error is defined using the percentage difference between the predicted revenue and the actual revenue. A positive value represents over-optimism.

A negative value represents over pessimism. For example, a prediction of $110,000 when actual revenue is $100,000 would be a forecast error of 10%. A prediction of $90,000 when actual revenue is $100,000 would be a forecast error of -10%. Among firms in the control group, please predict the average forecast error. For example, a value of "random" means you are forecasting that, on average, firms are over-optimistic by "random"% relative to actual sales. A value of "random" means you are forecasting that, on average, firms are over-pessimistic by "random"% relative to actual sales.

Firms were asked to report 5 scenarios AFTER the revenue forecasting. These scenarios were the revenue for their lowest, low, middle, high, and highest possible sales outcomes. Among firms in the control group, please predict the percentage of firms whose revenues were either below their lowest scenario, or above their highest scenario. For example, a firm that gave revenue scenarios of $1000, $2000, $2500, $3000 and $5000 would be outside that range if their actual realization came in as $400 but would be within the range if it came in at $4000. For example, a value of "random" means you are forecasting that "random"% of the firms in the control had actual revenues outside their lowest and highest scenarios.

In the previous question, you had forecasted the baseline win rate to be "random"%. Please predict the treatment effects for each intervention in percentage points. For example, a value of "random" represents a win rate of "random"% for that treatment group. Alternatively, a value of "random" represents a win rate of "random"% for that treatment group. You can hover over the treatment groups as a reminder of their definition.

Data Use - Firms are prompted to check a financial dashboard containing their recent revenues prior to making forecasts.

Incentives - Firms were offered up to $400 for accurate forecasts.

Training - Firms were trained in making forecasts for ten hypothetical businesses similar to the survey sample.

Scenario thinking - Firms were asked about their lowest, low, middle, high, and highest outcomes before the revenue forecasting, along with their likelihood.