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 for their financial support. We would like to thank seminar participants at the AEA, AFE, Chinese University of Hong Kong, Census Bureau, Chicago, Columbia, Davis, EMC, ERMAS, ESEM, HBS, NBER, Santa Clara, Stanford, UCSC, USC, UVA, WEAI, and Wharton. We would also like to thank Mert Akan, Patrick Collison, Eilin Francis, An He, Nate Hilger, Ana-Maria Mocanu, Alex Novet, Derek Ozkal, and Ethan Yeh for their feedback and support. AEARCT R-0007058. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Nicholas Bloom, Mihai A. Codreanu, and Robert A. Fletcher. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

ABSTRACT

We partner with a large US payment-processing company to run a 5-year, 10 wave panel survey of incentivized quarterly sales forecasts on over 6,000 firms. We match firm predictions to proprietary revenue data to assess accuracy. We find firms forecast poorly, with issues of inaccuracy, over-optimism, predictable errors and over-precision. To assess the causes of these forecasting issues we run experiments on: (i) data use, (ii) incentives, (iii) forecasting skill, and (iv) contingent thinking. We find greater data use primarily decreases noise and reduces over-precision, while higher incentives moderate over-optimism. Both moderately increase accuracy. The other two treatments have no impact. These results suggest forecasting biases can be reduced but are hard to eliminate. In a simple simulation model, we show these biases change firm responsiveness to changes in taxes and productivity, highlighting their macro importance.

Nicholas Bloom
Stanford University
Department of Economics
579 Jane Stanford Way
Stanford, CA 94305-6072
and NBER
nbloom@stanford.edu

Mihai A. Codreanu
Stanford University
Department of Economics
579 Jane Stanford Way
Stanford, CA 94305
mihaic@stanford.edu

Robert A. Fletcher
Instacart
robertsfletch@gmail.com

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

1. Introduction

Firms' expectations are a critical aspect of economics, playing a central role in theories ranging from Keynes's *General Theory* to workhorse models of investment and hiring under uncertainty. In much of the dynamic programming literature firms are modeled as forward-looking and rational. However, evidence for the rationality of firms' expectations is mixed, leading to widespread debate in both public discourse and academic literature.

Investigating firm expectations is challenging for two reasons. First, collecting realistic firm forecasts is difficult. In public forecasts, like company earnings guidance provided to financial markets, managers have incentives to bias forecasts to appeal to investors, customers or employees. Private surveys are usually unpaid, so managers may rush responses or provide responses they believe the surveyor wants (e.g. Bertrand and Mullainathan, 2001). Second, it can be hard to assess the accuracy of forecasts. Financial and administrative data can itself be biased to optimize share prices or tax payments. Assessing forecasts against follow-up survey data can also be problematic if, for example, managers who are over-optimistic in forecasting are also over-optimistic in reporting their sales to survey teams.

To address these challenges, this project took two steps.

First, we worked with a leading international financial services company, we call "FinTech", to collect detailed sales expectations from US businesses. Our panel has over 6,000 unique firms in ten waves from 2019 to 2024. To ensure engaged participation we paid respondents a $50 base fee, and a $25 fee per additional short survey completed. To elicit accurate responses respondents were also given an additional accuracy reward of up to $400 for forecasting their firm's next-quarter revenue within 10% of their actual revenue. To ensure accurate forecast assessments these predictions were evaluated using FinTech's transaction revenue data for these firms. While providing these rewards was expensive, with over $1,100,000 paid to managers, it yielded the first large scale, high-frequency incentivized panel of US firm forecasts.

Second, in waves five to ten of the survey we ran four randomized control trials to evaluate different drivers of forecast biases. First, we randomized firms' exposure to a data-dashboard on their recent FinTech revenue to investigate the extent to which historical sales data review reduces forecast biases. Second, we randomized accuracy rewards, ranging from $0 to $400, to evaluate to what extent greater incentives for accuracy can improve forecasting. Third, we randomized forecast training, providing 10 vignettes of a forecast training exercise to see to what extent improved forecasting skill can reduce bias. Finally, we randomized firms into evaluating scenarios in advance of making their forecasts, building on the "Superforecaster" literature[1] that suggests that considering extreme (highest and lowest) possible outcomes can improve forecasting.

Analyzing baseline firm forecasting data, we find four main types of biases. First, firms had inaccurate forecasts, with only 18% of forecasts within 10% of next-quarter sales (so eligible for the accuracy reward). This was so low that even using last quarter's revenue would have had a higher win rate by 1.4 percentage points. Second, firms were over-optimistic, forecasting

[1] See, for example, Gardner and Tetlock, 2015.

revenues that were 16.3% above realization. This optimism bias was pervasive, arising in every wave of the survey, including during the stable macro periods of 2019 and 2023, in the recession of 2020, and in the recoveries of 2021 and 2022. This bias arose in one-time respondents as well as repeat respondents, with no diminution in optimism bias with survey experience. Third, firms had predictable errors in that forecast errors were correlated with lagged forecast errors, and lagged outcomes like sales growth rates and levels. Finally, firms were over-precise. The survey elicited not only point-forecasts but also second-moment forecasts using 3-bin and 5-bin outcomes and probability questions. Firms provided subjective forecast variances that were approximately half the true variance.

We used the randomized control trials to identify possible causes of these forecasting biases.

We found the data dashboard exposure reduced inaccuracy, bias and over-precision. Those treated to the dashboard took just below 45 seconds more to complete the survey, decreased forecast noise by just below 15% and as a result increased their win rate by 2 percentage points. This suggests even short data exposures can improve forecasting performance. Surprisingly, despite the dashboard being free to access, firms that were randomized into the dashboard in a current wave did not show improved forecast accuracy (or dashboard use) in future waves. This suggests firms were over-confident in their forecasting ability, under-appreciating the value of additional information (e.g Moore and Healy, 2008).

The reward randomization reduced over-optimism bias. Firms randomized into the highest $400 reward bucket saw this optimism drop by around 70%. As a result, firms increased their win rates by 0.9 percentage points for each additional $100 dollars in rewards. This implies firms' optimism is pliable – firms are aware of it, and if adequately incentivized, can reduce their bias. One explanation is that being over-optimistic helps self-motivate managers to work harder, but if provided with sufficiently generous rewards they can be more realistic, a form of motivated reasoning (e.g Benabou and Tirole 2002).

Finally, neither the forecast training nor the contingent thinking interventions had any impact on forecast accuracy, bias, error predictability or over-precision. This is despite both performance in the training and in the contingent thinking scenario questions being associated with higher forecast accuracy in the cross-section. Overall, this suggests firms forecasting biases do not arise from a lack of forecast training or consideration.

The results suggest that these biases – inaccuracy, over-optimism, predictable errors and over-precision – are not merely superficial mistakes that disappear with basic training or nudges. Rather, they appear to be deep-seated behaviors that firms can only partially mitigate through targeted incentives or data usage. So, to investigate their impact we develop an illustrative simulation model of firms with these forecast biases. We show these biases not only reduce firm profitability and alter firm-size, but can also have macro implications, changing firms' dynamic response to tax and productivity shocks. Consequently, understanding and addressing these "built-in" biases could prove important for both firms and policymakers.

The paper relates to four literatures.

First, several papers explore firm expectations about macroeconomic variables, including Coibion and Gorodnichenko (2015), Coibion et al. (2018 and 2020), Tanaka et al. (2019) and Farmer, Nakamura and Steinsson (2024). In these models forecasting errors are typically modelled as arising due to sticky information (e.g. Mankiw and Reis 2002) or noisy information (Sims 2003 and Woodford 2003). Another line of the literature argues these forecasting biases in professional firms and analysts are due to representativeness (e.g. Bordalo et al., 2019 and 2020) where agents focus excessively on the most recent or salient information.

Second, there is a literature evaluating the biases of managers. Earlier work includes Malmendier and Tate (2005, 2008), who show that almost half of CEOs behave in a way consistent with over-optimism, most notably by holding excessive levels of their own company's stock. Until recently, few studies examined managerial beliefs directly (see, e.g., Manski, 2018; Shleifer, 2019) or considered quantitative models featuring beliefs. Exceptions include Bachmann and Elstner (2015), who studied German firms using directional rather than point-estimate forecasts, and Gennaioli et al. (2016), who examine CFO earnings forecasts showing their forecasting power for investment and predictable errors.[2] Ma et al. (2024) use a sample of Italian firms to show that sales forecasts are unbiased but conditionally biased, while investment appears to show under-responsiveness to sales forecasts. Another finding is over-precision. Ben-David, Graham, and Harvey (2013) examine forecasts from a similar set of executives, revealing executives provide excessively narrow forecast bands. Barrero (2022) also finds that managers have overly narrow subjective forecast distributions and make predictable extrapolative forecast errors.

Our research also contributes to the literature using randomized control trials for assessing and mitigating these biases (e.g. List, 2011). This literature is limited in using firms or firm managers as experimental group, potentially due to the high costs associated with and eliciting material effort from them, but there is a growing use of such experiments in accounting and finance (Floyd and List, 2016).

Third, a large literature has documented psychological biases that challenge rationality assumptions (e.g. Santos-Pinto and Sobel, 2005, Moore and Healy, 2008, Malmendier and Tate 2015 or Benabou and Tirole 2016). In the case of company managers, these may be emerge in equilibrium due to potential firm and individual benefits including increased firm innovation, as measured by patents and patents citations (Galasso and Simcoe, 2011; Hirshleifer, Low and Teoh 2012), or higher likelihood of being promoted (Goel and Thakor, 2008). Nonetheless, other papers highlight costs, including overpayment for target companies, seeing external funding as too costly (Malmendier and Tate, 2005), aggressive corporate investment (Ben-David, Graham, and Harvey, 2013) and underinvestment in information production (Goel and Thakor, 2008).

Recent work shows that persistent overconfidence can arise from biased memory, whereby managers selectively recall prior successes over failures (e.g. Huffman, Raymond and Shvets, 2022). At the theoretical level, Brunnermeier and Parker (2005) argue that decision makers may choose optimistic beliefs for increasing own utility, even when such beliefs lead to systematic forecast errors. More broadly, overconfidence, overoptimism, representativeness, and other

[2] See also Puri and Robinson (2007), Graham, Harvey, and Puri (2013, 2015), Bachmann et al. (2018)

biases that arise when individuals make preferences are central themes in behavioral finance (Barberis and Thaler, 2003).

Finally, our paper contributes to the literature investigating the aggregate effect of information frictions in managerial expectations. David et al. (2016) develops a steady-state production model to quantify improvements in aggregate efficiency resulting from a well-developed stock market. David and Venkateswaran (2019) employ a similar model to investigate the effect of information frictions—managers imperfectly observing future productivity—on aggregate efficiency, while Ma et al. (2024) model the micro and macro impact of the forecast biases they uncover in Italian firms.

Section 2 explains our survey design, Section 3 discusses our forecast biases while Section 4 discusses the results of our randomized control trials. Section 5 examines the impact of our firms' forecasting biases on aggregate outcomes, while Section 6 concludes.

2. Data

2.1. Survey Design and Details

We partnered with a large international payments technology operating in the United States, henceforth referred to as FinTech. They operate both online and in person payments systems, with some examples shown in Figure 1. The baseline sample of firms was constructed from their universe of US businesses with active accounts. Businesses had to be for-profits, and the contact emails had to be individuals or titles (i.e., they could not consist of phrases such as info@, admin@, or contact@). We sent 10 waves of surveys over 5 years from 2019 to 2024, with the sample for wave 1 in spring 2019 consisting of 18,000 randomly chosen firms, stratified to over-sample larger companies. Surveys were sent out about every 6 months to allow time between survey waves to assess firms' 3-month survey forecasts and pay any rewards before the next wave. 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 35,000 additional firms.

The surveys were targeted at business founders, obtaining a sample of approximately 95% of responses coming from an original founder or current CEO. The survey response rate was 15% on average across waves. We found that finishing the survey was close to random, with only a weak correlation with observables including size, revenue, and industry (R^2 of 0.01).

Each survey wave consisted of a core set of questions on the finances 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.

Firms were generally contacted with an invitation e-mail and three follow-ups each spaced five days apart. Firms were given $50 in Amazon vouchers to respond to the first wave of the survey

and then $25 for each subsequent wave. In addition, they were also given a $25 per survey wave accuracy bonus for their next quarter revenue forecast if these came within 10% of their realized revenue. This $25 accuracy reward was randomized from $0 to $400 as part of our interventions in waves 5, 7 and 9. Across all 10 waves of the survey we received just over 18,000 valid responses. We paid out around $1,100,000 of forecast payments and accuracy rewards. More details of the data structure can be found in the Data Appendix.

2.2. Business Characteristics

The characteristics of the firms and CEO are reported in Table 2. The firms have a mean revenue of just over $750,000 a year, with 6 employees. They are relatively online focused with 70% of their revenue online and just below 60% of total revenue via FinTech.

The firms span the entire United States with coverage across almost all states and industries (Figure 2), although with a skew towards industries like software, business services, travel and clothing which have a higher online representation. The average firm completed 3 survey waves, with 106 firms completing all 10 waves. The CEOs are typically 39 years old, with 41% female. They are also well educated: 95% have some college, and 31% have a Masters, JD, MBA or PhD

2.3. Payment Data

Using the FinTech data we can precisely match firms with their aggregate revenue, total transactions and average transaction value. This is valuable in allowing us to accurately assess high frequency firm responses against actual revenue data. This information is also readily available to firms via their FinTech account dashboard. We do not use tax, accounting, administrative or self-reported survey measures of incomes to assess forecast accuracy to avoid concerns over noise or bias in these data sources.

3. 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. We explicitly limited the predictions to FinTech revenue so that we could directly verify their forecasts. Because of the different transaction fees, timing and customer types firms track revenues by channel.[3] The surveys were designed to be short, and the median duration was 16 minutes, with the follow-ups averaging 10 minutes.

In each survey, respondents were promised an Amazon gift card if their three-month prediction was within 10% of their actual revenue. Prior to the fifth wave this reward was a fixed value of $25 to ensure that respondents had a financial incentive to provide credible and considered responses. In waves 5, 7, and 9, we randomized the reward amounts offered to respondents (see Appendix Figure A1 for screenshots). Specifically, 25% of respondents were randomly assigned to receive no reward, another 25% continued with the base amount of $25, and the remaining 50% were randomly placed into eight equally sized groups (6.25% each) corresponding to

[3] For example, cash payments have no payments fees while some online vendors charge up to 30% for providing payment infrastructure.

reward values of $50, $100, $150, $200, $250, $300, $350, and $400. The reward amounts were independently randomized in each wave and independently of all 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 the probabilities of these such that they add to 100% (see question in Appendix Figure A2). 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[4] to constrain the range to [-2,2] to avoid large values for near-zeros in predicted or actual revenue:

$$Forecast\ Error_{it} = \frac{(Predicted_{it} - Actual_{it})}{(0.5 * Predicted_{it} + 0.5 * Actual_{it})}$$

3.1 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 have to face adjustment costs or make (partially) irreversible or decisions—such as investment, hiring, pricing, or marketing—they need to forecast future demand.[5] In our data we see that firms with greater forecast accuracy are significantly larger, faster growing, more likely to survive and have a higher management score (Appendix Figure A3). 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, 16.8% of companies said forecasting was a top or top 3 factor in these business decisions, and 64.5% one of many factors (Appendix Figure A4). Firms stated that forecasting is most important in setting materials expenditures, followed by advertising. Interestingly, around three quarters of firms said that forecasting is one of the top or top 3 factors for at least one of these decisions. Hence, forecasting accuracy appears important for performance.

3.2 Forecasting Issues

We identify four key forecasting issues which we discuss in turn.

Forecasting Inaccuracy: The first issue is the low accuracy of forecasts in our sample of 18,060 predictions. As Table 2 shows the mean squared forecast error is 0.812, highlighting a large number of large forecast errors. Perhaps most notably, only 18% of firms are within 10% of their next quarter's sales and are rewarded with the accuracy bonus. This forecasting accuracy rate is not only low in absolute terms but is also low compared to other simple forecasting metrics. For

[4] Adapted from the Davis and Haltiwanger (1992) growth rate measure.
[5] Some classics in this literature include Hayashi (1982), Shapiro (1986), Caballero and Engel (1993), 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.

example, using their prior quarters sales as their forecast yields a higher win rate in the baseline pre-experimental (waves 1 to 4) sample by 1.4 percentage points. Given firms should have a large amount of additional information over and above their last quarter revenue – such as imminent orders, new products, seasonality and customer flow – they should be able to do substantially better.

Bias: The average forecast error is +16.3%. The presence of over-optimism bias has been shown before in the literature, for example as surveyed in Manski (2018). However, these results often come from uncompensated surveys benchmarked against subsequent survey responses, so it is possible either the original forecast or the follow-on survey data may be inducing the bias. Other evidence comes from public forecasts, often by professionals, where other incentives (e.g., appeal to investors) may bias predictions.

Table 3 shows that this over-optimistic bias is a persistent feature across the business cycle. A unique particularity of our large, high-frequency panel sample of incentivized sales forecasts benchmarked directly against sales data is the long temporal span. By chance the survey period spanned from 2019—a period of stable expansion—through the deep COVID-19 recession of 2020, the rapid recovery of 2021 and 2022, 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). We also examined a group of persistent responders—firms that made 5 or more responses—who would have repeatedly been paid the survey response fee and (if they won) the accuracy bonus. This would make our payments particularly salient for this group since they would have repeatedly received them in prior survey waves. This sample also shows a highly significant over-optimism bias of around 8% for this persistent respondent group, which is most visible in the expansion periods.

Predictable Errors: The firms' forecast errors are predictable. As Figure 3 shows, current forecast errors are extremely highly correlated with lagged forecast errors. 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[6] between each 3-month survey wave. This makes this autocorrelation of forecast errors even more striking. One explanation is heterogeneous bias. Another as highlighted in Coibion and Gorodnichenko (2015 and 2018) is slow information updating due to rational inattention or information collection costs (e.g. Mankiw and Reis, 2002). This has to be long-lasting as we see in Table 4 that firms' forecast errors are correlated with up to 5 lags of prior forecast errors stretching back 2.5 years[7]. Moreover, as we explore further in the experimental section, firms do not appear to use easily available information in their current forecasts, including their own historic revenue data.

Over-precision: We also evaluate the higher moments of firm forecasts by eliciting their forecast distributions. We asked firms to report their 3-bin (lowest, middle and highest) forecasts or 5-bin (lowest, low, middle, high, and highest) forecasts for revenue in the next three months. We combined the forecasted states with their perceived probabilities of each of these outcomes with a requirement these sum to 100% (see Appendix Figure A2 screenshot). We convert these

[6] The exact timing between when surveys were run was mainly influenced by logistical constraints related to the FinTech's collaboration and gift card acquisition and processing.
[7] Longer lags have positive (but declining) coefficients, but samples sizes are too small for longer lags to be statistically significant.

into continuous distributions by converting the point probabilities to intervals using the mid-points between bins, following Manski (2003) and Altig et al. (2020).

Armed with this subjective probability distribution, we can evaluate their properties by examining the distribution of the realizations for the entire sample of firms. One way to do this is to use the large numbers of firms responding in each cross-section and determine at what point in a firm's subjective distribution their actual sales fall in the following quarter. If our firm panel provided accurate subjective expectations, we would expect to see 1% of firms falling in the bottom 1% of their subjective distribution, 1% in the second 1%, 1% in the third 1%, and so on.[8]

In Figure 4, we plot the actual distribution of firms' quarterly sales on the FinTech platform compared to their subjective forecasts. An accurate forecast would have 1% of the sample in each of the 1% percentile bins, as indicated by the flat red line. Instead, we find that 13.2% of outcomes are below the lowest bin, 15.3% are above the highest bin, and on average only 0.55% of outcomes are in each percentile of the central bins from 40% to 60%. Thus, firms are providing subjective expectations that are too precise, resulting in their realized sales outcomes too frequently falling outside the predicted bins (and too infrequently within their predicted range). One measure of over-precision we will use later is the share of forecast outcomes outside the cumulative distribution function of the subjective distribution; in Figure 4, this is just below 30%. Another measure is to compare, as in Table 2, the variance of their subjective forecast distribution, which is 0.272, to the variance of forecast errors, which for this sample of 9,905 observations was 0.775, implying their forecast variance is just above 35% of their realized variance.

Qualitatively, this over-precision result is similar to results found in Ben-David et al (2013) and Barrero (2022) sample, albeit on a different sample of firms with no incentivization. It also shadows the results discussed in Moore and Healey (2008) that individuals find it particularly hard to make accurate forecasts on second moments, tending to be over-precise.

4. Experiments

There are several potential explanations for the forecasting issues we have uncovered in our firms. One is informational, in that perhaps firms are making optimal forecasts but with stale or incomplete information. This could be due to rational inattention or noisy signals (e.g. Coibion and Gorodnichenko 2015) or diagnostic expectations (e.g. Bordalo et al. 2020). Another explanation is that firms may be optimizing a different objective function from forecast accuracy. For example, perhaps for self-motivation reasons they are deliberately over-optimistic, which can explain the large pliable positive bias. 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.

[8] In line with the literature (e.g. Altig et al. 2020; Manski, 2018), we first transform the subjective distribution into growth rates and then split the distribution in 3 or 5 intervals (depending on the exact question elicitation procedure in the relevant survey wave), each with their reported probabilities.

To evaluate these and other possible explanations, we ran four experiments on our firms, randomizing data access, financial incentives, training and scenario evaluations. We discuss these four experiments in turn, ordering them in the sequence we ran them chronologically.[9]

4.1. **Improving Data: Randomizing Usage of the Data Dashboard**

The first experiment we tried was to improve firms' information set. As part of using FinTech to process their payments, businesses are given access to a financial dashboard. This dashboard, (shown in appendix Figure A5) appears when users open their account. The dashboard allows users to see various financial metrics, including past revenue.

One motivation for an information treatment was that while the average forecast error was a striking 16.3%, in a simple decomposition, the bias term accounts for less than 5% share of the mean squared forecast error across the waves (mean squared forecast error across all waves was 0.812).[10] This aligns with the Kahneman et al. (2021) focus on the importance of noise in prediction errors, suggesting perhaps more data could improve forecasting.

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% 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) their exact revenue over the last three months from their dashboard.[11] The control group was only asked to report their revenue from the last three months with no suggestion to use their dashboard. Note that even for the control group the dashboard is easily accessible on the top page when opening the firm's FinTech account.

In Table 5 we report on the results of the dashboard experiment. In the top-panel 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 45 seconds, on a non-treated dependent variable mean of 1184 seconds. This indicates that the dashboard viewing took approximately three-quarters of a minute. In column (2) we report the arc-percentage error on the last three months' reported revenue, which shows an extremely large 9.5 points reduction, representing around 21% of the baseline non-treated mean. This shows, perhaps unsurprisingly, that viewing the dashboard improves the accuracy of reporting their *prior quarter* revenue. In column (3) we report the impact on the time taken on the prediction question, which increases by 8.2 seconds on a mean of 72 seconds, suggesting having additional data only moderately increases how long they take

[9] The first two experiments - the dashboard and rewards experiments - were preregistered on 1/21/2021. This was after the experiment began but before the quarterly results data was available to evaluate the first wave. The other two experiments were added in later waves, so were not preregistered, although neither had any significant impact.

[10] Note that $E(error^2) = E(error)^2 + E(error - \big(E(error)\big)^2$ so Mean Squared Error = Bias^2 + Noise. At baseline, the bias^2 is 0.027, which is 3.3% of the mean squared error of 0.812.

[11] 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.

to make their revenue forecasts. Finally, column (4) reports the share of digits that are rounded in their forecast as an indicator of forecast accuracy. This is calculated as the share of "0" trailing digits in their forecast.[12] We see the dashboard lowered rounding by approximately 2.2 percentage points.

Panel B examines forecasting outcomes. In column (5) we see the win rate rises by 2% on a base of 17.2% in the control group. Hence, survey participants randomized into the dashboard intervention won about 12% 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. The mean prize value in the dashboard waves was $83.1, and given the prizes were in the form of Amazon gift vouchers this likely represents a post-tax value. This yields an estimated post-tax return of $133 per hour.[13] This likely represents a very conservative estimate of the perceived return, as most respondents vastly over-estimate their likelihood of winning (detailed below).

This value of dashboard time is even more notable because the 45 seconds involved two tasks – finding their firm ID number and finding their last quarter revenue. Finding their firm ID is more onerous, requiring multiple clicks, while revenue is displayed on the dashboard home screen. So, firms could probably look up recent revenue data in perhaps 20 seconds, pushing up the value of dashboard review time to about $300 per hour after tax, equivalent to perhaps a pre-tax salary of $500 or $600 an hour for the typical higher tax-rate manager.

Perhaps even more strikingly, despite this beneficial impact of dashboard randomization on winnings there is no dynamic effect. Dashboard randomization has no impact on future dashboard use or future win-rates (see Table A1). So, even after managers are randomized into using the dashboard, increasing their current win rate, they do not increase dashboard in future waves.

Column (6) provides another measure of accuracy – the squared errors – finding that the mean-squared error fell by 9.7 points on a base of 67.6%, so a reduction of 14.3%. While this is not a massive decrease, it indicates the dashboard intervention improves the forecast accuracy by reducing forecast noise (outliers) rather than reducing forecast bias. Indeed, figure 5 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 most of the higher percentiles for forecast error (that were positive). Hence, the dashboard information compressed the squared forecast error by reducing the tail forecast errors.[14]

In column (7) we examine the mean forecast-error and find this drops by 3.7 percentage points, suggesting more information also has a moderate impact on reducing over-optimism (decreasing

[12] So, for example, a revenue forecast of $62,500 would be 40% (two fifths of the digits are trailing "0"), while $650,000 would be 66%, and $603,105 would be 0% (none of the trailing digits are "0").

[13] Calculated as 0.02*83.1*3600/45

[14] Interestingly, despite being a much smaller percentage of the baseline error, we cannot reject that the treatment effect in the squared historic error and in the squared forecast errors is the same in absolute value. Given that the results suggest that this is the main mechanisms through which firms increased their win rates and the results of the quantile regression showing over-optimists may have been impacted more, this may be connected to Huffman, Raymond and Shvets (2022) models of motivated beliefs in which individuals are motivated to distort memories of feedback and preserve unrealistic expectations.

it by about a third from the control group baseline). In column (8) 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. Column (8) reports this interaction, showing the dashboard led firms to reduce their current forecast error coefficient on their lagged forecast error by 0.05. While this is insignificant, the coefficient magnitude is about 18.5% of the mean value on lagged error of 0.269, providing some weak evidence that better information also reduces error predictability.[15]

Finally, Table 5 Panel B column (9) examines the impact of the dashboard experiment on the extent of over-precision, measured as the share of the forecasts outside of the cumulative distribution function of the subjective distribution. We see this fell by 2.9%, or almost a 10% reduction given the base of 29.9%. This highlights how the information intervention not only increased firms point-estimate precision but also improved the higher-moments of their subjective forecast distributions, even though these estimates were not directly incentivized.

Collectively, the results of the dashboard intervention support models of inadequate data use by firms, perhaps because firms are excessively confident in their forecasting ability or have motivated beliefs distorting past performance. Firms can substantially improve their win rates, earning $300 an hour from dashboard use to improve their accuracy. But they fail to do so, even in future waves after being randomized into the dashboard intervention in the current wave. This excessive confidence in their own forecasting ability aligns with the results from section 3.3, 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 64.2%, more than three-times their actual win rate of 20% in that wave.[16] This aligns with a long literature on over-precision (related to "over-confidence" and "over-placement") which highlights how individuals tend to over-estimate the accuracy of their forecasts and under-estimate the value of additional information (e.g. Moore and Healy, 2008).

4.2. **Altering Attention: Randomizing Forecast Accuracy Rewards**

We also tested randomizing the reward for forecast accuracy. The motivation for this was concerns over how seriously firms were taking the survey question. We thought offering up to $400 for an accurate survey would elicit more considered and accurate responses.

In waves 5, 7 and 9 we randomized 50% 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% 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% chance of having no reward at all. A further random 25% stayed with

[15] Note the issues with a smaller sample here. We can only examine firms subject to the dashboard RCT who have also made forecasts in a prior wave.

[16] 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%*".

the base amount of $25, and the remaining 50% 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[17].

In Table 6, Panel (A) we first examine the survey response statistics. In column (1) we see the randomized incentives increased total survey time by 6.3 seconds per $100 of additional reward, albeit not statistically significant. In column (2) we see the randomization had no impact on the prior quarter's revenue reporting. In column (3) we see the randomization increased the time taken to complete the forecasting question by 16.1 seconds per $100 reward. So, a firm offered the maximum $400 reward would have taken about 1 minute longer to complete the survey on a base of 51.8 seconds (for those in the control group), a substantial increase in response time. This suggests the reward did induce firms to take more time considering their forecasts. 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%.

In Table 6 Panel (B) we first see in column (5) that the win rate increased by 0.9% per $100 of additional rewards. Firms offered the maximum $400 of reward improved their forecasting accuracy by 3.6% on average. However, even for these firms randomized into the maximum reward amount their win-rate increased by only around 22% on the base win-rate of 16.1%. In column (6) we see this rewards intervention also had no impact on the squared error.

In column (7) we see the rewards intervention did cut the mean error (bias) by 1.9 points per $100. So, firms offered the maximum reward of $400 cut their bias by 7.6 points, reducing total bias by 69.7% (on the base bias of 10.9%). In Figure 6 we estimate quantile regressions for the impact of $100 of additional reward for percentiles from 5% to 95% and see uniform reduction in the forecast level, suggesting that, unlike the dashboard experiment, the whole distribution of observed biases has shifted down. Finally, in columns (8) and (9) we see the reward had no significant impact on either the error predictability or the degree of over-precision.

These results collectively suggest that providing stronger financial rewards for forecast accuracy induces firms to spend more time making their forecasts. This leads them to cut their over-optimism, with the maximum reward cutting almost three quarters of this bias. This is the main mechanism that leads to the moderate increase in the win rate, as the reward did not shift accuracy as measured by the squared forecast error.

This impact on reducing the mean bias is consistent with a model in which firms are aware of their optimism bias. When offered a large financial incentive for forecast accuracy they manage to temper this, suggesting this is a pliable bias. Perhaps they find over-optimism is helpful for motivating themselves, employees and/or investors, but a material financial reward tamps this down, a form of motivated reasoning (e.g. Kunda 1990 and Benabou and Tirole 2002).

[17] We chose the maximum reward of $400 a year 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 $575 (three survey fixed payments of $75, two regular wins of $25 and one win of up to $400). In practice no respondent randomized into the $400 rewards correctly forecasted three times consecutively, so even this $575 bound was never met.

One interesting comparison is the return per second of time spent predicting in the dashboard compared to the reward experiments. We calculate each additional prediction second increased the win rate by 0.055% in the incentives experiment compared to 0.045% in the dashboard experiment, which is remarkably similar. However, the mechanism through which these two experiments operated is different – the incentives mainly operate through decreasing bias and the dashboard by decreasing noise.

4.3. **Training: Randomizing Forecast Training**

After analyzing the results from waves 5 and 6 showing the moderate impacts of the data intervention and marginal impact of 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. Randomization was independent across waves and independent of other interventions.

The training exercise asked firms to make forecasts for a series of 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 7 left panel). They were then shown a suggested forecast we created (e.g. Figure 7, right panel), which placed approximately ½ weight on the last quarter revenue and 1/6 on the prior three quarters revenues, based on an econometric best-fit regression of sales revenues for our sample frame.[18] This process was repeated for ten hypothetical firm vignettes, with the vignette ordering randomized by firm. Figure 8 contains three more example vignettes.

Before discussing the impact of the forecasting exercise, we present some results validating this training exercise. We start by examining the progress of firms forecasting accuracy over the course of the ten 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% 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. Second, firms with higher average forecast accuracy across the training exercise had higher forecast accuracy on *prior forecasts* they made in previous incentivized survey waves for their firms.[19] So, the firms' performance on this forecasting exercise was highly correlated to their own forecasting ability in prior waves.

Collectively these two results – that training improved firms forecasting on the training exercise, and that this training exercise was related to their prior forecasting ability – was auspicious for

[18] 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.
[19] In a regression of the arc-percentage error of firms actual forecast error on their average forecast error in the training exercise we have a coefficient (standard-error) of .302 (0.066).

the training increasing win rates. As we see, much to our surprise, this forecast training appeared to have no impact on improving firm forecasting accuracy or win rates.

In Table 7 we first present the survey and forecast diligence metrics. In column (1) we see managers randomized to the forecasting exercise took 45 seconds longer to complete the survey. In column (2) 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 (3) we see the training increased the prediction response time by 4.5 seconds, although this was insignificant. In column (4) we see it had no material impact on rounding, albeit some suggestive negative impact.

In Table 7 panel B we examine the impact of the forecast training of their forecast accuracy. In column (5) we see it has no impact on the win rate. Surprisingly it also may have increased the forecast squared error in column (6), perhaps due to survey fatigue. In columns (7), (8) and (9) we see the training also had no impact on the mean forecast error (bias), over predictability or over-precision.

Collectively, these results suggest that while forecast training improved firms' forecasting skills, it did not improve their own performance forecasting their next quarter revenue. This appears to rule out that the firm's poor performance was due to challenges in understanding the survey question or making forecasts.

4.4. **Scenarios: Randomizing Consideration of Extreme Scenarios**

After the null result from the interventions in waves 5 to 7, in wave 8 we experimented with 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 and 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 changes in tax and interest rates, macroeconomic events, 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, 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. By asking them to consider scenarios from their lowest to higher outcomes it may reduce forecast tail errors.

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

uncertainty from their scenario questions versus the 3-month later realized squared forecast error on the y-axis. We see a highly significant relationship. 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 forecast accuracy.

Table 8 panel (A) starts by evaluating the survey diligence metrics. We see this exercise has a negative impact on overall survey duration in column (1), by decreasing it by 53.4 seconds. Most of this decrease seems to come from reducing the time taken in the prediction exercise, 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 (2). Finally, in column (4) we see it has no impact on the forecast rounding. In Table 8 panel (B) we find across all five metrics – win-rate, squared error, bias, 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.

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.

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 under use available data. 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. 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. Overall, both experiments only have small relative effects, showing that these biases are hard to moderate.

5. Modelling aggregate implications

5.1. Baseline model

In sections 3 and 4 we found evidence that firms have four biases: (a) inaccuracy, (b) over-optimism, (c) predictable errors, and (d) over-precision. In this section, we build a simple model to illustrate the potential impacts of these biases at the micro and macro level.

We create an economy with a population of firms using Cobb-Douglas production with just a capital input, $Y = AK^{\alpha}$. Productivity, A, is stochastic with a log-autoregressive function. Firms also pay taxes, τ, so after tax revenue is $(1-\tau)pAK^{\alpha}$. Taxes are stochastic and follow a two-state Markov process.

Forecasting matters because capital has an adjustment cost in the form of a 10% resale cost. Capital also depreciates at 10% a year. Firms profit maximize by choosing their capital stock K' following the Bellman equation, where r is the interest rate:

$$V(K,A) = \max_{K'}(1-\tau)pAK^{\alpha} - (r+\delta)K - C(K,K') + \frac{1}{1+r}E(V(K',A'))$$

5.2. Forecasting Biases

To simulate the impact of the empirical forecasting biases we highlight in section (3) we compare the baseline model to five different models. One each with the individual forecasting issues individually, and one with all combined. These four individual biases are:

Inaccuracy: firms make capital decisions believing their productivity, A_t, is equal to it's true value plus a noise term with zero mean and a quarter of the variance of A_t.

Over-optimism: firms forecast their productivity to be a quarter of the standard deviation of A_t higher than its actual distribution.

Persistent errors: firms forecast with weight 0.9 on their true current productivity A_t and 0.1 on the unconditional mean of the productivity distribution $\bar{A}$.

Over Precision: firms forecast using half of the true standard deviation of the productivity process.

We simulate an economy with 5,000 firms operating on a quarterly frequency (with all parameters reported in Table A.2). In Table 9 we outline the impact of these forecasting biases on some key moments of the simulated economy, reported as percentage deviations from the no-forecast bias baseline simulation. In column (1) we see all these biases unsurprisingly reduce profits. In columns (2) and (3) we see inaccuracy increases output due to the convexity of optimal capital[20] as does over-optimism and over precision. Auto-correlated errors reduce output as it reduces volatility, which due to convexity lowers output. This impact of persistent errors dominates in the combined forecast error case, lowering total output. Finally, looking at volatility in columns (4) and (5) we see inaccuracy and over-precision increase this by inducing more responsiveness from firms, with correlated errors reducing this for the same reason in reverse.

In Figure 10 we also show these forecasting errors have material impacts on the response of firms to tax and productivity shocks. We simulate the aggregate impact of a Period 1 rise in the tax rate from 20% to 25% (left panel) and a 10% drop in productivity (right panel). We see large negative impacts on capital, output and profit. Importantly, there is also a wide spread in response rates due to variations in the forecast bias, ranging up to 50%. The largest response to these shocks is typically for over optimistic firms and the smallest response for firms with persistent errors, although even this ranking is dependent on the outcome and type of bias.

These simulations highlight two important results. First, these forecasting biases can have large material impacts. Firms are forward looking, and if their expectations are altered from the true-driving process this impacts the entire path of inputs, outputs and profits. Second, if firms have heterogeneous forecasting biases, so for example some firms are over-optimistic and others have persistent errors, this will lead to substantial heterogeneity in the firm-level response to common

[20] Note in the flexible capital case that the optimal capital stock, $K^* \propto A^{\alpha/(1-\alpha)}$, is convex in A.

shocks. As such, forecasting biases can have major implications for micro and macro outcomes, as well as for optimal policy choices.

6. Conclusions

We partner with a large US payment-processing company to run a 5 year, 10 wave panel survey of incentivized quarterly sales forecasts for over 6,000 firms. We match firm predictions to proprietary revenue data to assess forecasting accuracy.

We found four main types of biases. First, firms had inaccurate revenue forecasts, with only 18% of forecasts within 10% of next-quarter sales. Second, firms were over-optimistic, forecasting revenues that averaged 16.3% above realizations. This optimism bias was pervasive, arising in every wave of the survey, including during the stable macro periods of 2019 and 2023, in the recession of 2020, and in the recoveries of 2021 and 2022. Third, firms had predictable errors in that forecast errors were correlated with lagged forecast errors, and lagged outcomes like sales growth rates and levels. Finally, firms were over-precise, providing subjective forecast variances that were below half of the true variance.

We used the randomized control trials to identify possible causes of these forecasting biases.

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, firms that were randomized into the dashboard in a current wave did not show improved forecast accuracy or dashboard use in future waves. This suggests firms were 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 reduced over-optimism by almost three quarters. 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.

Both the rewards and dashboard experiments increased the win rates, but the effects were less than a quarter of the baseline, even for the firms offered the highest reward amounts. These impacts are particularly small since at the baseline only 18% of firms forecasted their sales within 10% of the realized value. The training and scenario experiments had no effects.

The results suggest that these biases - over-optimism, predictable errors and over-precision – are innate, and can be reduced but are hard to fully eliminate. To investigate their impact we develop a simulation model of firms with these forecast biases. We show these biases not only reduce firm profitability and alter firm-size, but they also have macro implications, changing firms' dynamic response to tax and productivity shocks.

We think these results pave the way for future research. First, understanding the motives for maintaining biased forecasts. Second, designing and testing alternative interventions that leverage improved data analysis, using machine learning predictive tools, and incorporating behavioral insights may be more effective. Finally, embedding these forecasting behaviors into general equilibrium frameworks would identify broader macroeconomic implications and inform policy interventions aimed at improving resource allocation under uncertainty.

Bibliography

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

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

Bachmann, Rüdiger, Kai Carstensen, Stefan Lautenbacher, and Martin Schneider. "Uncertainty and Change: Survey Evidence of Firms Subjective Beliefs," 2018, mimeo.

Bachmann, Rüdiger, and Steffen Elstner. "Firm Optimism and Pessimism." *European Economic Review* 79 (October 1, 2015): 297–325.

Barberis, Nicholas, and Richard H. Thaler. 2003. "A Survey of Behavioral Finance." In *Handbook of the Economics of Finance*, (edited by Constantinides, Harris, and Stulz).

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

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. "Managerial Miscalibration*." *The Quarterly Journal of Economics* 128, no. 4 (November 1, 2013): 1547–84.

Benabou, Roland and Jean Tirole, "Self confidence and Personal Motivation", *Quarterly Journal of Economics,* 2002, 117(3), 871-915.

Benabou, Roland and Jean Tirole, "Mindful economics: the production, consumption and value of beliefs", *Journal of Economic Perspectives*, 2016, 141-164.

Bertrand, Marianne and Sendhil Mullainathan, "Do people mean what they say? Implications for subjective survey data". American Economic Review, vol 91(2), pp 67-72.

Bertola, Giuseppe, and Ricardo J. Caballero. "Irreversibility and aggregate investment." *The Review of Economic Studies* 61.2 (1994): 223-246.

Bloom, Nicholas, et al. "Really uncertain business cycles." *Econometrica* 86.3 (2018): 1031-1065.

Bordalo, Pedro, et al. "Diagnostic expectations and stock returns." *The Journal of Finance* 74.6 (2019): 2839-2874.

Bordalo, Pedro, N. Gennaioli, Yueran Ma, and A. Shleifer, "Overreaction in Macroeconomic Expectations", 2020, American Economic Review, (2020), 2748-2782.

Bradshaw, Mark T., et al. "The Interactive Role of Difficulty and Incentives in Explaining the Annual Earnings Forecast Walkdown." The Accounting Review, vol. 91, no. 4, 2016, pp. 995–1021.

Bruhin, Adrian, Luís Santos-Pinto, David Staubli, "How do beliefs about skill affect risky decisions?," 2018, Journal of Economic Behavior & Organization, Volume 150, Pages 350-371

Brunnermeier, Markus, K., and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review*, 95 (4): 1092–1118**.**

Caballero, R and E. Engel (1993): "Microeconomic Adjustment Hazards and Aggregate Dynamics," *Quarterly Journal of Economics*, 108, 359–383.

Coibion, O and Y. Gorodnichenko "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," American Economic Review, 2015, 105 (8), 2644–78.

Coibion, O., Y. Gorodnichenko, and S. Kumar (2018): "How do firms form their expectations? new survey evidence," American Economic Review, 108, 2671–2713.

Coibion, O., Y. Gorodnichenko, and T. Ropele (2020): "Inflation expectations and firm decisions: New causal evidence," Quarterly Journal of Economics, 135, 165–219.

Cooper, Russell W., and John C. Haltiwanger. "On the nature of capital adjustment costs." *The Review of Economic Studies* 73.3 (2006): 611-633.
David, Joel, Hugo Hopenhayn, and Venky Venkateswaran, "Information, Misallocation, and Aggregate Productivity," Quarterly Journal of Economics, 2016, 131, 943–1005..
David, Joel M and Venky Venkateswaran, "The Sources of Capital Misallocation," American Economic Review, 2019, 109 (7), 2531–67.
Davis, S., and J. Haltiwanger (1992): "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," *Quarterly Journal of Economics*, 107, 819–863.
Dixit, A., and R. Pindyck (1994): *Investment Under Uncertainty*. Princeton, NJ: Princeton University Press.
Epley, Nicholas and Thomas Gilovich, "The mechanics of motivated reasoning", *Journal of Economic Perspectives,* 2016, pp 133-140.
Farmer, Leland, Emi Nakamura and Jon Steinsson, "Learning about the long-run", 2024, *Journal of Political Economy,* October, 132 (10).
Floyd, Eric, and John A. List. "Using Field Experiments in Accounting and Finance." SSRN Working Paper 2016
Galasso, Alberto and Timothy S. Simcoe, "CEO Overconfidence and Innovation", Management Science 2011 57:8, 1469-1484
Gennaioli, N., Y. Ma, and A. Shleifer (2016): "Expectations and investment," NBER Macroeconomics Annual, 30, 379–431.
Goel, A.M. and Thakor, A.V. (2008), Overconfidence, CEO Selection, and Corporate Governance. The Journal of Finance, 63: 2737-2784.
Graham, J. R., C. R. Harvey, and M. Puri (2013): "Managerial attitudes and corporate actions," Journal of Financial Economics, 109, 103–121.
Graham, J. R., C. R. Harvey, and M. Puri (2015): "Capital allocation and delegation of decision-making authority within firms," Journal of Financial Economics, 115, 449–470.
Haran, Uriel and Don Moore, "A better way to forecast", *California Management Review,* 2014, 57(1).
Hayashi, F. (1982): "Tobin's Marginal Q and Average Q: A Neoclassical Interpretation," *Econometrica*, 50, 213–224.
Hirshleifer, D., Low, A. and Teoh, S.H. (2012), Are Overconfident CEOs Better Innovators?. The Journal of Finance, 67: 1457-1498.
Huffman, David, Collin Raymond, and Julia Shvets. 2022. "Persistent Overconfidence and Biased Memory: Evidence from Managers." American Economic Review, 112 (10): 3141–75.
Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. *Noise: A flaw in human judgment*. Hachette UK, 2021.
Kunda Z. The case for motivated reasoning. Psychol Bull. 1990 Nov;108(3):480-98. doi: 10.1037/0033-2909.108.3.480. PMID: 2270237.
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
Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar. "A Quantitative Analysis of Distortions in Managerial Forecasts." Cambridge, MA: National Bureau of Economic Research, March 2024 (update on 2020).
Malmendier, U. and G. Tate (2005): "CEO overconfidence and corporate investment," Journal of Finance, 60, 2661–2700

Malmendier, U. and G. Tate (2008): "Who makes acquisitions? CEO overconfidence and the market's reaction," Journal of Financial Economics, 89, 20–43.
Malmendier, Ulrike, and Geoffrey Tate. "Behavioral CEOs: The role of managerial overconfidence." *Journal of Economic Perspectives* 29.4 (2015): 37-60.
Mankiw, N. G., and R. Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." Q.J.E. 117:1295–328
Manski, Charles F. *Partial identification of probability distributions*. Springer Science & Business Media, 2003.
Manski, Charles F. "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise." *NBER Macroeconomics Annual* 32 (April 1, 2018): 411–71.
Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. Psychological Review, 115(2), 502-517
Moore, Don, and Uriel Haran. "A simple tool for making better forecasts." *Harvard Business Review* (2014): 2-4.
Moore, Don A., Elizabeth R. Tenney, and Uriel Haran. "Overprecision in judgment." *The Wiley Blackwell handbook of judgment and decision making* 2 (2015): 182-209.
Ropele, Tiziano, Yuriy Gorodnichenko, and Olivier Coibion, "Inflation Expectations and Misallocation of Resources: Evidence from Italy," AER: Insights, 2024.
Puri, Maju and David Robinson (2007), "Optimism and economic choice", *Journal of Financial Economics*, 86, pp 71-99.
Santos-Pinto, Luís, and Joel Sobel. 2005. "A Model of Positive Self-Image in Subjective Assessments." American Economic Review, 95 (5): 1386–1402.
Shapiro, M. (1986): "The Dynamic Demand for Labor and Capital," *Quarterly Journal of Economics*, 101, 513–542.
Shleifer, A. (2019): "The Return of Survey Expectations," NBER Reporter.
Sims, C. A. 2003. "Implications of Rational Inattention." J. Monetary Econ. 50:665–90.
Stokey, N., and R. Lucas, (1989): *Recursive Methods in Economic Dynamics, With E. Prescott*. Cambridge, MA: Harvard University Press.
Tanaka, Mari, Nicholas Bloom, Joel M. David, and Maiko Koga. "Firm Performance and Macro Forecast Accuracy." Journal of Monetary Economics, March 5, 2019.
Tetlock, Philip E.; Gardner, Dan (2015). *Superforecasting : the art and science of prediction*. New York: Crown Publishers.
Woodford, M. 2003. "Imperfect Common Knowledge and the Effects of Monetary Policy." In Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, edited by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, 25–58. Princeton, NJ: Princeton Univ. Press.

Data Appendix

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 2.1, 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 month a small round of 50 firms was send out first in the first month as a test sample. This checked for survey errors, bugs or other issues. The following month the main rounds 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. 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 month back or forwards. For example, one wave was delayed by a month when FinTech's Stata license expired (which was required for the sample creation).

For the data to appear in our final sample we drop those that did not provide a prediction, that had no transactions in the past 3 months (except for wave 10 where we could not enforce this due to data availability reasons) as they presumably left FinTech, firms that rushed through the survey (defined as finishing it in less than 7 minutes), and those that did not have, on average, at least 5% of their revenue on FinTech. The four experiments were cross-randomized, before sending the survey invites, according to the last digits of the account numbers (that were as good as random):

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 A5.

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% of respondents were randomly assigned $0, another 25% continued with the base amount of $25, and the remaining 50% were randomly placed into eight equally sized groups of $50, $100, $150, $200, $250, $300, $350, and $400. Example of the questions shown to firms are in Appendix Figure A1. These rewards 9as 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 7), and are then shown our own forecast. Due to ex-ante possible dynamic effects, we randomized the sample into three groups, each of a third: one third saw the training exercises before making the prediction, one third after (to maintain a roughly constant total survey length), and one third

did not see it at all. The results are shown for the group that saw the training before making the prediction compared to those that did not.

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 A2). 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.

For creating these probability distributions from these 3-point and 5-point forecasts we used the method in Altig et al. (2020). We transform each scenario's sales prediction into piecewise probability densities by first normalizing the difference from the previous three months (to keep the scale comparable across firms). Internal bins are then assumed to run from the mid-point between forecasts. For example, if the "lowest", "low", "middle", "high" and "highest" forecasts were 2, 4, 6, 10 and 15 then the "low" bin would be defined as running from 3 to 5 and the "middle" bin as 5 to 8 and the "high" bin from 8 to 12.5. To generate the lowest and highest bins we follow the "Manski bin bounds" approach which extend the low and high tails of each firm's distribution by 2 the midpoint differences between adjacent scenario points. This ensures realized outcomes are less likely to fall outside the elicited support altogether. In the example above that would mean the "lowest" bin would be from 0 to 3 and the highest best from 12.5 to 20. Probability masses are then uniformly spread across the bins.

We compute the mean forecast as a probability-weighted average across scenarios, derive the variance (and thus the coefficient of variation) to capture subjective uncertainty, and pinpoint each realized outcome's percentile on the predicted distribution. These measures also let us gauge over-precision (how often actual outcomes fall outside the firm's best-worst forecast ranges).

Table 1 – Project Timeline

Wave	Dates	Responses	Extra Module	Randomized Control Trial
1	Jan-Apr 2019	2785	Baseline	
2	May-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	Aug-Sep 2021	2760	Forecasting	Dashboard, Reward, Forecast Training
8	Apr-Aug 2022	1272	Forecasting	Dashboard, Scenarios
9	Sept 2022-Jan 2023	1735	Forecasting	Dashboard, Reward, Forecast Training
10	Nov 2023-Jan 2024	1208	Funding	Dashboard, Forecast Training, Scenarios

Notes: This table reports the timeline of the 10 waves of the Project (column 1), by the outer window dates of the start of the wave responses across all waves and rounds (column 2), and total number of clean survey responses (column 3), totaling 18,060. Each of these waves had a specific extra module (column 4), and starting wave 5, at least one randomized control trial for the incentivized 3-month sales forecasting questions (column 5).

Table 2 – Firm Descriptive Statistics

	Observations	Mean	Median	Std. Dev.
Panel A: Firms				
Revenue last 12 months ($'000s)	6582	750.96	71	2581.25
Number of Employees	6316	5.57	2	8.50
Revenue Online, %	6591	70.33	90	35.72
Revenue from FinTech, %	6594	59.51	61	31.21
Number Survey Waves	6594	2.74	2	2.31
Panel B: CEOs				
Age (years)	6121	39.36	38	10.53
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
Variance (Subjective Distribution)	9905	0.272	0	0.55

Notes: This table reports the descriptive statistics for 6,594 unique firms (totaling 18,060 responses) that have completed our survey, for firm, CEO, treatment and forecasts characteristics. The table includes information on the mean (column 2), the median (column 3), and the standard deviation (column 4), at entry. For the information in Panel B (CEOs), we have used information from the immediate next wave after entry if that was available. The variables shown are (in order): all firm earnings in the last 12 months, the number of employees (including full and part-time), percentage of the total revenue online, percentage of the total revenue on the FinTech company, and the number of survey waves with valid responses; age of the CEO, self-reported gender of the CEO, 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 (prediction of the 3-months forecasting question within 10% of realized value on FinTech), 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 calculated as the square of the arc-percent forecast errors, and variance coming from the 3- or 5-bin subjective distribution, calculated using the Altig et al. (2020) algorithm. All employee and age variables were winsorized at the 2.5^{th} and 97.5^{th} percentiles of the distribution in the first survey wave they responded.

Table 3 – Over-Optimism Over Time

Wave	Dates	Obs	Over-Optimism	Obs (Persistent Respondents)	Over-Optimism (Persistent Respondents)
1	Jan-Apr 2019	2785	0.35***	848	0.18***
2	May-Aug 2019	1685	0.22***	805	0.10***
3	Oct 2019-Jan 2020	1994	0.29***	974	0.15***
4	Apr-May 2020	1602	0.06***	974	-0.06**
5	Sep-Oct 2020	1765	0.08***	1036	0.03
6	Jan-Apr 2021	1254	0.08***	939	0.05**
7	Aug-Sep 2021	2760	0.05***	892	0.03
8	Apr-Aug 2022	1272	0.11***	590	0.08**
9	Sept 2022-Jan 2023	1735	0.04**	699	0.02
10	Nov 2023-Jan 2024	1208	0.27***	659	0.24***
All	**Jan 2019-Jan 2024**	**18060**	**0.16*** **	**8416**	**0.08*** **

Notes: This table reports the averages forecasting 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) for a sample of 18,060 total responses (column 3) and 8,416 responses of persistent respondents, defined as users that had 5 or more answers over the 10 waves of the survey (column 5). Stars ** represent significance at 5% level (i.e. p-value of mean being 0 is <0.05) *** represent significance at 1% level (i.e. p-value of mean being 0 is <0.01).

Table 4 – Predictable Errors

	(1) Forecast Error	(2) Forecast Error	(3) Forecast Error	(4) Forecast Error	(5) Forecast Error
$Forecast\ Error_{i,t-1}$	0.260***	0.208***	0.245***		
	(0.018)	(0.024)	(0.039)		
$Forecast\ Error_{i,t-2}$		0.178***	0.100***		
		(0.023)	(0.033)		
$Forecast\ Error_{i,t-3}$		0.080***	0.073**		
		(0.023)	(0.035)		
$Forecast\ Error_{i,t-4}$			0.093***		
			(0.033)		
$Forecast\ Error_{i,t-5}$			0.089***		
			(0.031)		
$Lagged\ Growth$				-0.047***	
				(0.017)	
$Lagged\ Log(1 + Income)$					-0.045***
					(0.005)
Wave by Round FE	Yes	Yes	Yes	Yes	Yes
F-test (equal lags)		0.001	0.006		
R-squared	0.075	0.113	0.147	0.017	0.031
Observations	9108	3594	1460	9108	9108

Notes: The dependent variable is the forecast error 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). In columns (1)-(3) the independent variables are the lags of the forecast error, calculated from the same formula (for the relevant 3 months of forecasting) from the last 1 to 5 response waves where valid responses were recorded, in column (4), the growth rate in the last 3 months (which also uses arc-percentile growth, calculated as the 2*(Sales Last 3 Months-Sales Last 4-6 Months)/(Sales Last 3 Months+Sales Last 4-6 Months)), and in column (5) the lagged sales in the last 3 months, calculated as the log(Sales Last 3 Months+1). All sample was restricted to firms that had non-missing forecast errors, first lag of forecast errors, last quarter growth and lagged log(1+income), using the definitions above. All regressions include wave by round fixed effects. Stars ** represent significance at 5% level (i.e. p-value of mean being 0 is <0.05) *** represent significance at 1% level (i.e. p-value of mean being 0 is <0.01).

Table 5 –The Dashboard Treatment

Panel A: Survey and Forecast Diligence				
	(1) Survey Time (secs)	(2) Last 3 Months (Historic Error)^2	(3) Prediction Time (secs)	(4) Rounding (% digits)
$Dashboard_{i,t}$	44.718***	-0.095***	8.243**	-2.203***
	(9.632)	(0.020)	(3.206)	(0.449)
R-squared	0.013	0.017	0.029	0.008
N	9994	9881	5837	9994
Dependent Mean	1184	0.453	71.5	58.3

Panel B: Forecasting Outcomes					
	(5) Win Rate	(6) Forecast Error^2	(7) Forecast Error	(8) Error Predictability	(9) Over-Precision
$Dashboard_{i,t}$	0.020**	-0.097***	-0.037**	-0.050	-0.029***
	(0.008)	(0.022)	(0.016)	(0.038)	(0.009)
R-squared	0.006	0.023	0.013	0.067	0.032
N	9994	9994	9994	5594	9598
Dep. Mean	0.172	0.676	0.110	0.269	0.299

Notes: This table reports results of the dashboard treatment on survey and forecasting diligence outcomes (in Panel A) and main sales forecasting outcomes (in Panel B). The independent variable is a dummy taking the value 1 if a firm is treated by the dashboard, and 0 otherwise. In Panel A, the dependent variables are: survey time, calculated as the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); historic sales squared error, calculated as the [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 2); prediction time, calculated as the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 1200 seconds (column 3); rounding percentage of digits, calculated as the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panel B: win rate percentage, where winners are individuals who have finished the survey and their forecasts were within 10% of the realized value (column 5); forecasting 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 6); forecasting error squared which is calculated as the squared value of the 3-month forecasting error (column 7); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment with the dependent mean being the linear forecast lag coefficient for scaling purposes (column 8); and probability to have the actual 3-months outcome in the next 3 months outside of the predicted best and worst 3-months scenarios using Altig et al. (2020) subjective distribution calculation algorithm (column 9). 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 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) ** represent significance at 5% level (i.e. p-value is <0.05) *** represent significance at 1% level (i.e. p-value is <0.01).

Table 6 –The Rewards Treatment

Panel A: Survey and Forecast Diligence				
	(1) Survey Time (secs)	(2) Last 3 Months (Historic Error)^2	(3) Prediction Time (secs)	(4) Rounding (% digits)
Reward, $\$100s_{i,t}$	6.321	0.008	16.102***	-1.853***
	(4.686)	(0.009)	(1.902)	(0.216)
R-squared	0.008	0.004	0.036	0.015
N	6260	6215	3357	6260
Dependent Mean	1191	0.362	51.8	62.3

Panel B: Forecasting Outcomes					
	(5) Win Rate (%)	(6) Forecast Error^2	(7) Forecast Error	(8) Error Predictability	(9) Over-Precision
Reward, $\$100s_{i,t}$	0.009**	-0.009	-0.019***	0.010	0.002
	(0.004)	(0.010)	(0.007)	(0.019)	(0.004)
R-squared	0.005	0.005	0.005	0.069	0.006
N	6260	6260	6260	2668	6126
Dependent Mean	0.161	0.610	0.109	0.226	0.252

Notes: This table reports results of the rewards treatment on survey and forecasting diligence outcomes (in Panel A) and main sales forecasting outcomes (in Panel B). The independent variable is a variable taking the value of the rewards offered to the firm (in hundreds of dollars). In Panel A, the dependent variables are: survey time, calculated as the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); historic sales squared error, calculated as the [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 2); prediction time, calculated as the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 1200 seconds (column 3); rounding percentage of digits, calculated as the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panel B: win rate percentage, where winners are individuals who have finished the survey and their forecasts were within 10% of the realized value (column 5); forecasting 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 6); forecasting error squared which is calculated as the squared value of the 3-month forecasting error (column 7); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment with the dependent mean being the linear forecast lag coefficient for scaling purposes (column 8); and probability to have the actual 3-months outcome in the next 3 months outside of the predicted best and worst 3-months scenarios using Altig et al. (2020) subjective distribution calculation algorithm (column 9). 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 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) ** represent significance at 5% level (i.e. p-value is <0.05) *** represent significance at 1% level (i.e. p-value is <0.01).

Table 7 –The Training Treatment

Panel A: Survey and Forecast Diligence				
	(1) Survey Time (secs)	(2) Last 3 Months (Historic Error)^2	(3) Prediction Time (secs)	(4) Rounding (% digits)
$Training_{i,t}$	45.125***	0.074**	4.570	-0.634
	(13.165)	(0.030)	(3.864)	(0.621)
R-squared	0.009	0.020	0.027	0.007
N	5703	5606	4565	5703
Dependent Mean	1215	0.398	79.1	57.5

Panel B: Forecasting Outcomes					
	(5) Win Rate	(6) Forecast Error^2	(7) Forecast Error	(8) Error Predictability	(9) Over-Precision
$Training_{i,t}$	-0.014	0.060*	-0.009	0.072	0.016
	(0.011)	(0.032)	(0.023)	(0.064)	(0.013)
R-squared	0.006	0.031	0.018	0.083	0.039
N	5703	5703	5703	2494	5423
Dependent Mean	0.180	0.630	0.096	0.261	0.277

Notes: This table reports results of the training treatment on survey and forecasting diligence outcomes (in Panel A) and main sales forecasting outcomes (in Panel B). The independent variable takes value 1 if a firm sees training exercises before forecasting, and 0 otherwise (after/not at all). In Panel A, the dependent variables are: survey time, calculated as the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); historic sales squared error, calculated as the [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 2); prediction time, calculated as the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 1200 seconds (column 3); rounding percentage of digits, calculated as the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panel B: win rate percentage, where winners are individuals who have finished the survey and their forecasts were within 10% of the realized value (column 5); forecasting 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 6); forecasting error squared which is calculated as the squared value of the 3-month forecasting error (column 7); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment with the dependent mean being the linear forecast lag coefficient for scaling purposes (column 8); and probability to have the actual 3-months outcome in the next 3 months outside of the predicted best and worst 3-months scenarios using Altig et al. (2020) subjective distribution calculation algorithm (column 9). 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 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) ** represent significance at 5% level (i.e. p-value is <0.05) *** represent significance at 1% level (i.e. p-value is <0.01).

Table 8 –The Scenarios Treatment

Panel A: Survey and Forecast Diligence

	(1) Survey Time (secs)	(2) Last 3 Months (Historic Error)^2	(3) Prediction Time (secs)	(4) Rounding (% digits)
$Scenarios_{i,t}$	-53.423***	0.066	-32.094***	-0.461
	(19.326)	(0.047)	(3.947)	(0.903)
R-squared	0.015	0.025	0.030	0.009
N	2480	2412	2480	2480
Dependent Mean	1232	0.470	71.4	56.1

Panel B: Forecasting Outcomes

	(5) Win Rate	(6) Forecast Error^2	(7) Forecast Error	(8) Error Predictability	(9) Over-Precision
$Scenarios_{i,t}$	-0.006	0.005	0.022	0.007	0.032
	(0.016)	(0.050)	(0.034)	(0.066)	(0.021)
R-squared	0.007	0.039	0.018	0.061	0.011
N	2480	2480	2480	1931	2239
Dependent Mean	0.198	0.759	0.175	0.256	0.395

Notes: This table reports results of the scenarios treatment on survey and forecasting diligence outcomes (in Panel A) and main sales forecasting outcomes (in Panel B). The independent variable is a dummy taking the value 1 if a firm is treated by the scenarios, and 0 otherwise. In Panel A, the dependent variables are: survey time, calculated as the number of seconds spent on the survey since started, winsorized at 30 minutes (column 1); historic sales squared error, calculated as the [2*(Reported Sales Last 3 Months-Actual Sales Last 3 Months)/(Reported Sales Last 3 Months+Actual Sales Last 3 Months)]^2 (column 2); prediction time, calculated as the number of seconds spent on the forecasting the sales in the next 3-months page, winsorized at 1200 seconds (column 3); rounding percentage of digits, calculated as the number of trailing digits that are zero compared to the total number of digits in the 3-months forecast (column 4). In Panel B: win rate percentage, where winners are individuals who have finished the survey and their forecasts were within 10% of the realized value (column 5); forecasting 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 6); forecasting error squared which is calculated as the squared value of the 3-month forecasting error (column 7); error predictability which reports the interaction between the treatment and lagged error in a regression that also includes the lagged error and the treatment with the dependent mean being the linear forecast lag coefficient for scaling purposes (column 8); and probability to have the actual 3-months outcome in the next 3 months outside of the predicted best and worst 3-months scenarios using Altig et al. (2020) subjective distribution calculation algorithm (column 9). 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 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) ** represent significance at 5% level (i.e. p-value is <0.05) *** represent significance at 1% level (i.e. p-value is <0.01).

Table 9 –Impact of Forecasting Errors in the Simulation vs Baseline

Model	Average Profit (1)	Average Output (2)	Average Capital (3)	Output Volatility (4)	Capital Volatility (5)
Inaccuracy	-1.40%	+8.04%	+10.20%	+6.67%	+8.63%
Over-optimism	-14.56%	+42.87%	+56.47%	+55.70%	+76.40%
Auto-correlated errors	-3.00%	-15.16%	-17.90%	-44.71%	-55.27%
Over-precision	-1.26%	+10.05%	+12.54%	+8.99%	+10.69%
All forecast errors	-3.74%	+10.42%	+13.99%	-21.00%	-29.49%

Notes: This table reports the simulation results from simulating 1,000 firms' responses under different forecasting error scenarios, as described in the model in Section 5.1. The table reports results scenarios include inaccuracy, over-optimism, persistent errors, over-precision, and combined biases. Metrics include average output, average capital, output and capital volatility, coefficient of variation (CV), productivity-capital correlation, and average profits. Deviations are all relative to the baseline scenario (normalized to 100%). Parameters used for the model are presented in Table A2.

Table A1 – Dynamic Impact of Experiments on Firm Outcomes

	(1) Win Rate	(2) Forecast Error^2	(3) Forecast Error	(4) Error Predictability	(5) Over-Precision
Panel A: Dashboard					
$Dashboard_{i,t-1}$	-0.009	-0.034	-0.025	0.028	-0.002
	(0.012)	(0.029)	(0.022)	(0.050)	(0.014)
N	4585	4585	4585	3153	4358
Dependent Mean	0.212	0.561	0.081	0.536	0.299
Panel B: Rewards					
Reward, $\$100s_{i,t-1}$	-0.005	-0.021	-0.005	0.027	-0.004
	(0.006)	(0.014)	(0.010)	(0.024)	(0.007)
N	2926	2926	2926	1702	2736
Dependent Mean	0.211	0.527	0.077	0.550	0.252
Panel C: Training					
$Training_{i,t-1}$	0.013	-0.014	-0.020	0.025	0.020
	(0.019)	(0.049)	(0.036)	(0.093)	(0.024)
N	1931	1931	1931	1024	1761
Dependent Mean	0.203	0.591	0.125	0.618	0.277
Panel D: Scenarios					
$Scenarios_{i,t-1}$	0.011	-0.029	-0.026	0.051	0.018
	(0.029)	(0.068)	(0.051)	(0.053)	(0.031)
N	795	795	795	711	783
Dependent Mean	0.199	0.519	0.019	0.513	0.395

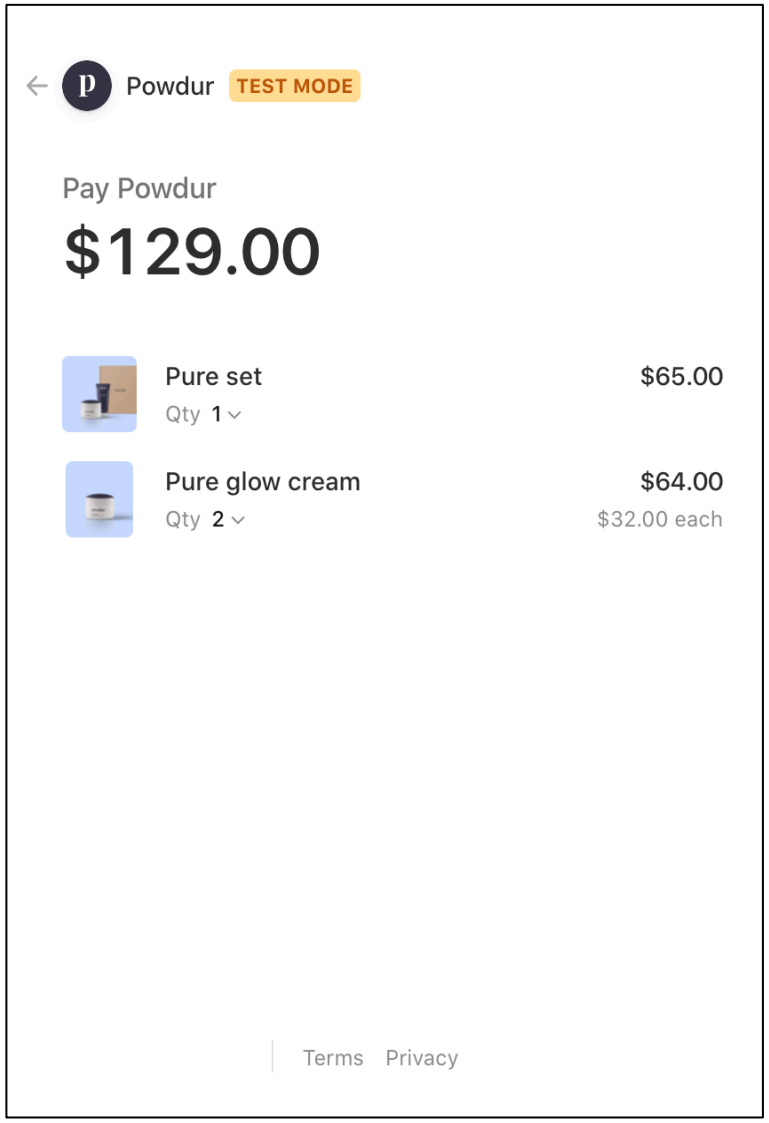
Notes: This table reports the dynamic impacts of the dashboard (Panel A), rewards (Panel B), training (Panel C) and scenarios (Panel D) treatments in the previous waves on forecast-related outcomes. Dependent variables are: win rate percentage, calculated as the win rate among respondents, where winners are individuals who have finished the survey and their forecasts were within 10% of the realized value (column 1); forecasting 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 2); forecasting error squared (or mean squared error) which is calculated as the squared value of the 3-month forecasting error defined above (column 3); error predictability which reports the interaction between the treatment and lagged error (2 periods ago) in a regression that also includes the lagged error (2 periods) and the treatment (so both linear terms) with the dependent mean being the linear coefficient of the lagged (2 periods) forecast error for scaling purposes (column 4); and probability to have the actual outcome of the forecast for the sales in the next 3 months outside of the predicted best and worst 3-months sales scenarios using Altig et al. (2020) subjective distribution calculation algorithm (column 5). All regressions also have reported the number of observations (when the dependent variables are defined, and for all the periods these are available). The control dependent variable mean is reported as the dependent variable mean for the untreated individuals in the survey waves where the treatments were run and all the independent and dependent variables where defined. 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) ** represent significance at 5% level (i.e. p-value of mean being 0 is <0.05) *** represent significance at 1% level (i.e. p-value of mean being 0 is <0.01).

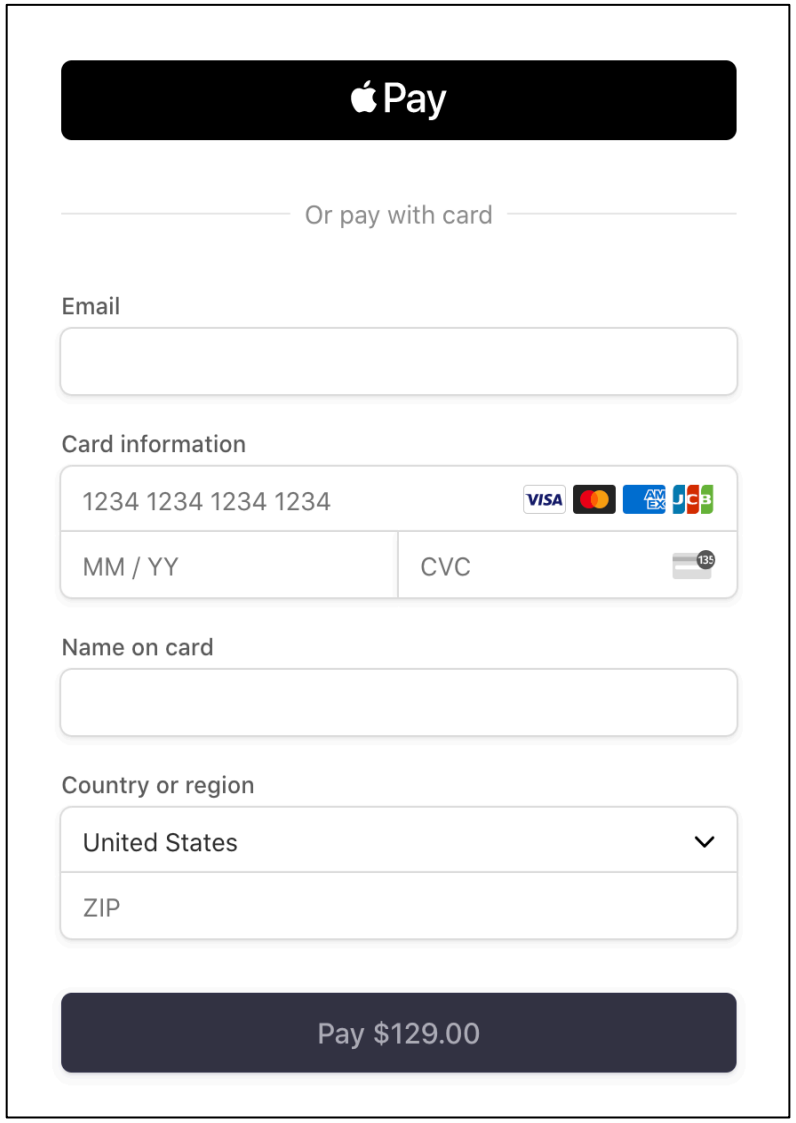
Table A2 – Model Parameters

Parameter	Symbol	Value	Description
Capital elasticity	α	0.8	Elasticity of output with respect to capital in production function
Adjustment cost parameter	Φ	0.1	Parameter for partial irreversibility in adjustment costs
Interest rate	r	2.5% per quarter	Quarterly interest rate (approx. 10% annually)
Tax rates	τ	0.20 and 0.25	Tax rates used in simulation
Discount factor	β	0.975	Quarterly discount factor (approx. 10% annually)
Productivity persistence	ρ	0.95	Persistence parameter in AR(1) productivity process
Productivity shock std. dev.	σ	0.1	Standard deviation of productivity shocks
Mean productivity	$\bar{A}$	1	Mean of productivity process
Depreciation rate	δ	2.59% per quarter	Quarterly depreciation rate (approx. 10% annually)
Tax transition matrix	P_t	[0.95, 0.05; 0.05, 0.95]	Tax rate transition probabilities, with 5% chance of tax rate change each quarter

Figure 1: We Partnered with a Fintech which Processes Payments Online and in Person

Examples of the types of online (left two) and in-person (right) payment options used by the FinTech's customers

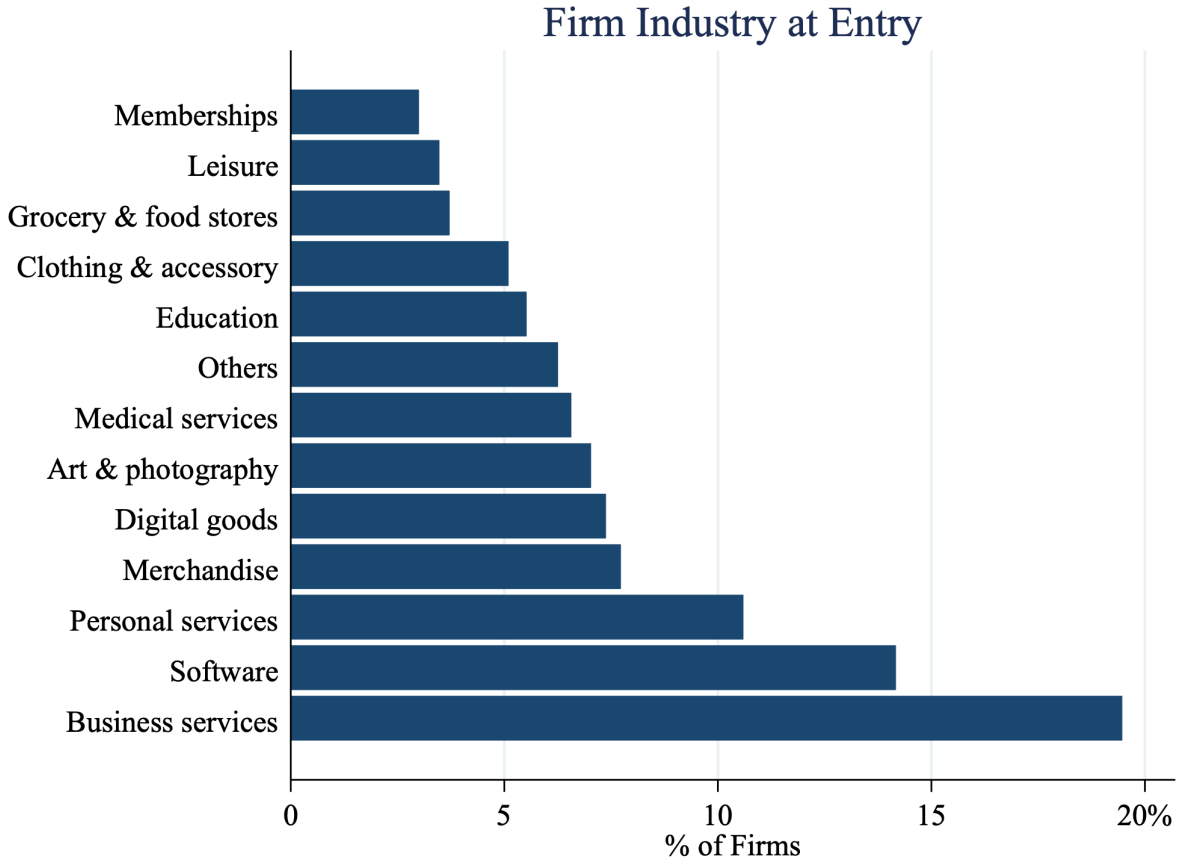


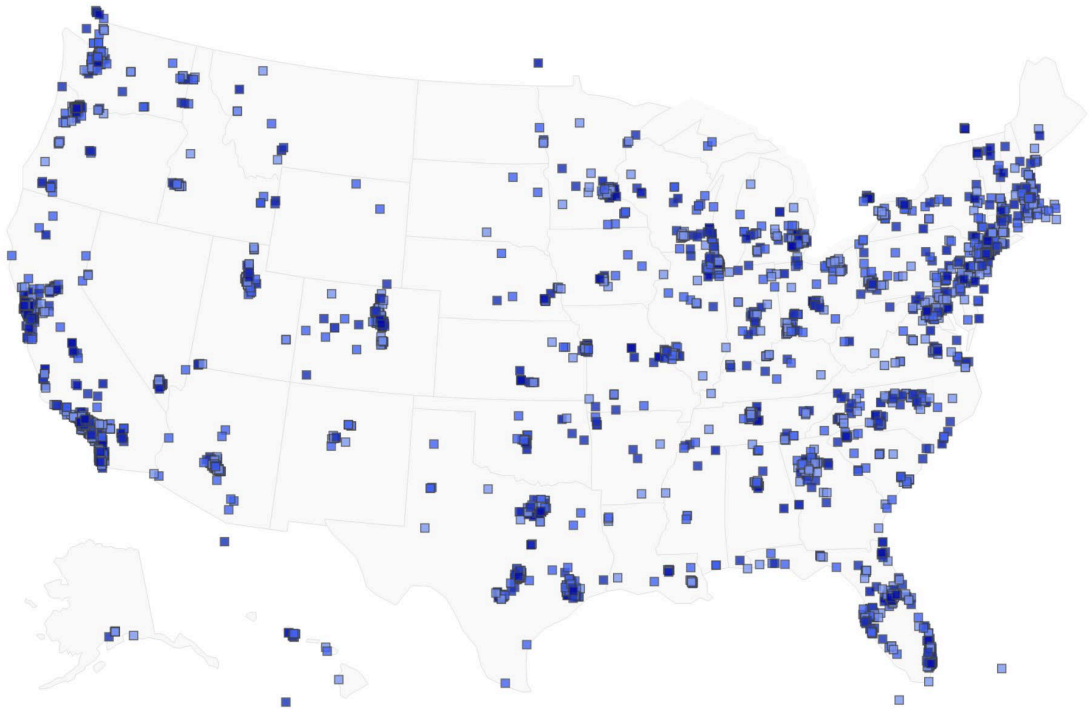




Note: the name of the company has been covered or blurred for anonymity. The right photo was taken by a co-author in a local gym.

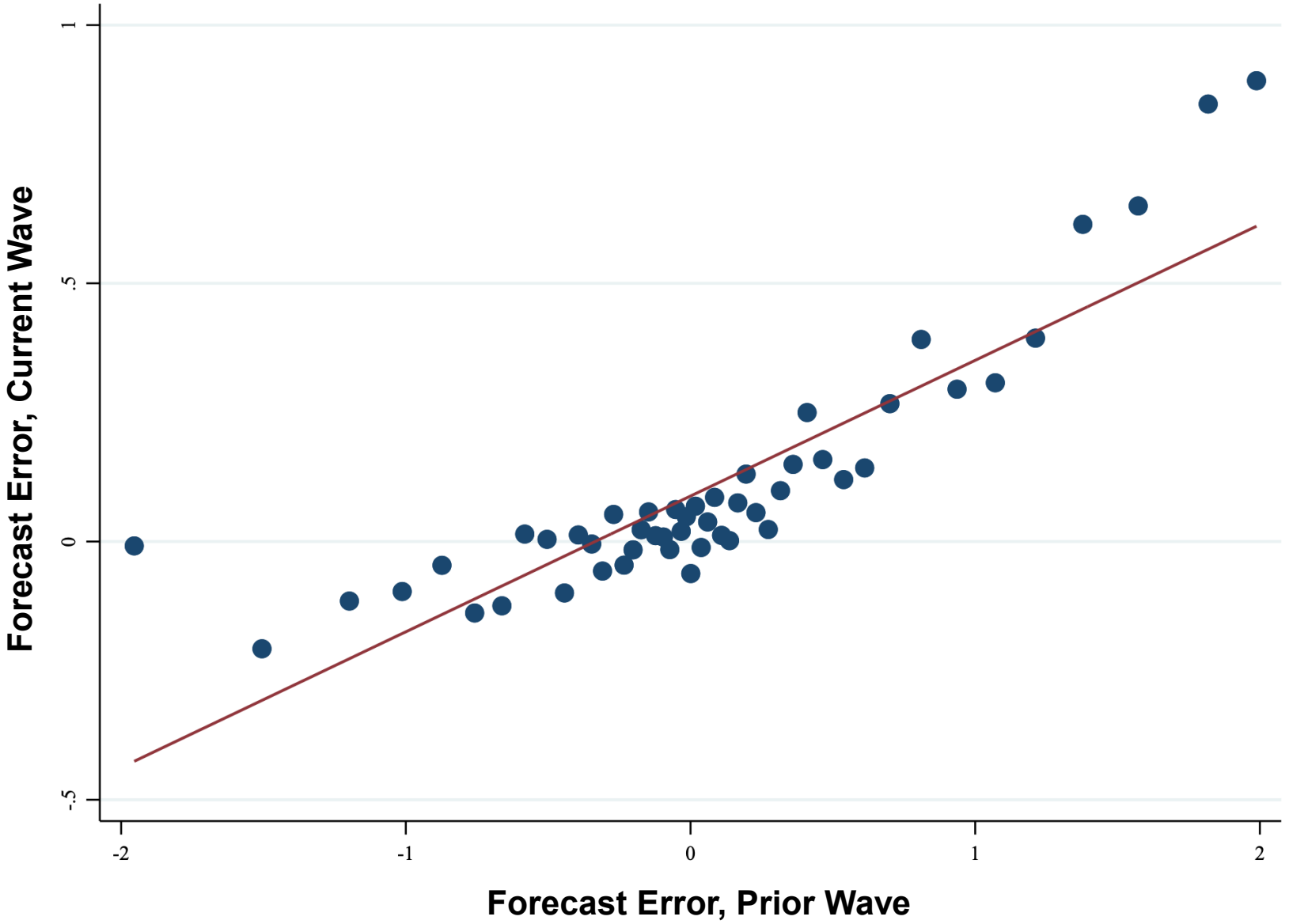
Figure 2: Broad Industry and Region Spread of Firms





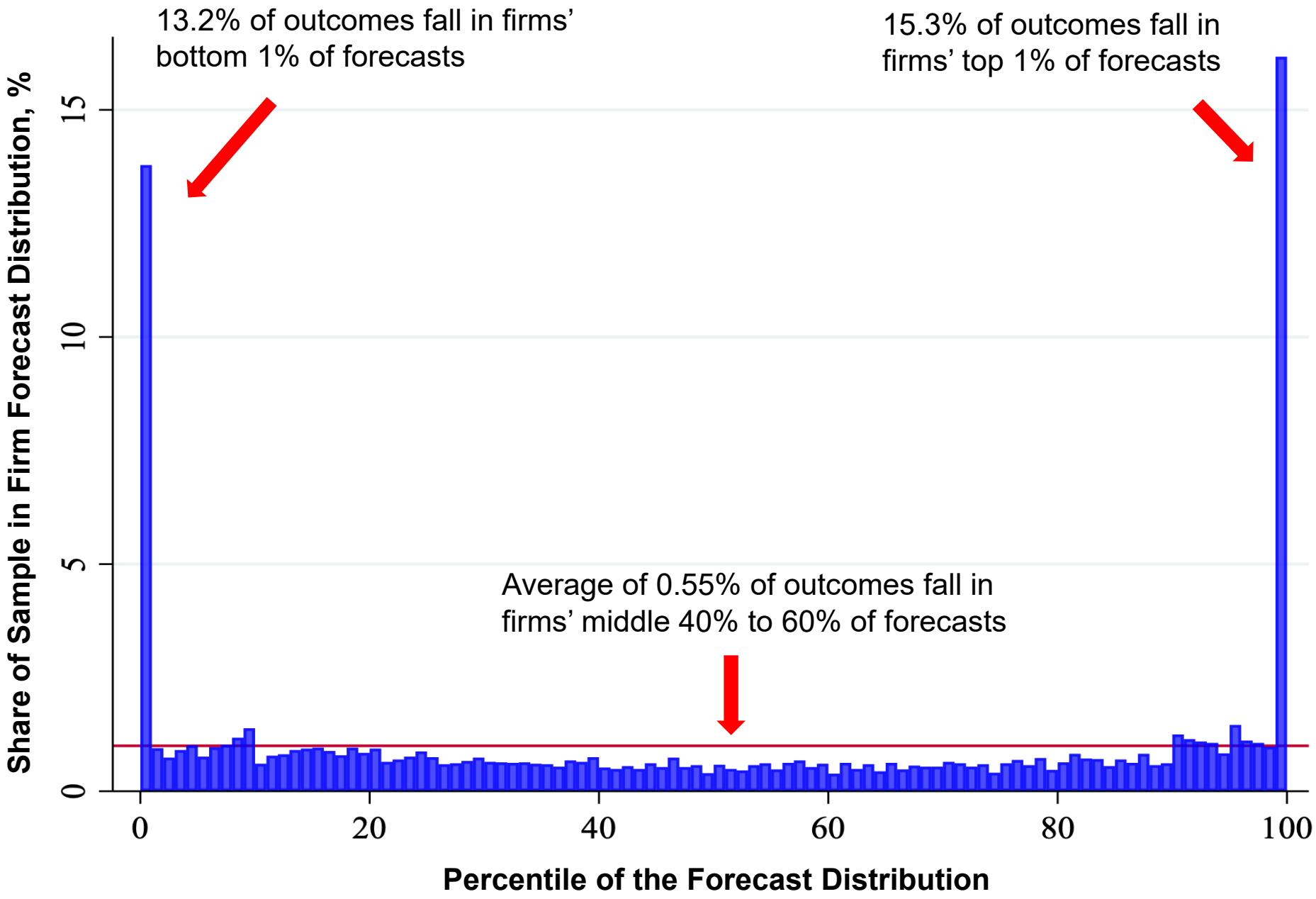
Notes: Left plot shows the histogram of firms by industry while the right plot shows the location of a random 10% sample of firms sampled, at the entry in our survey.

Figure 3: Forecast Errors are Predictable



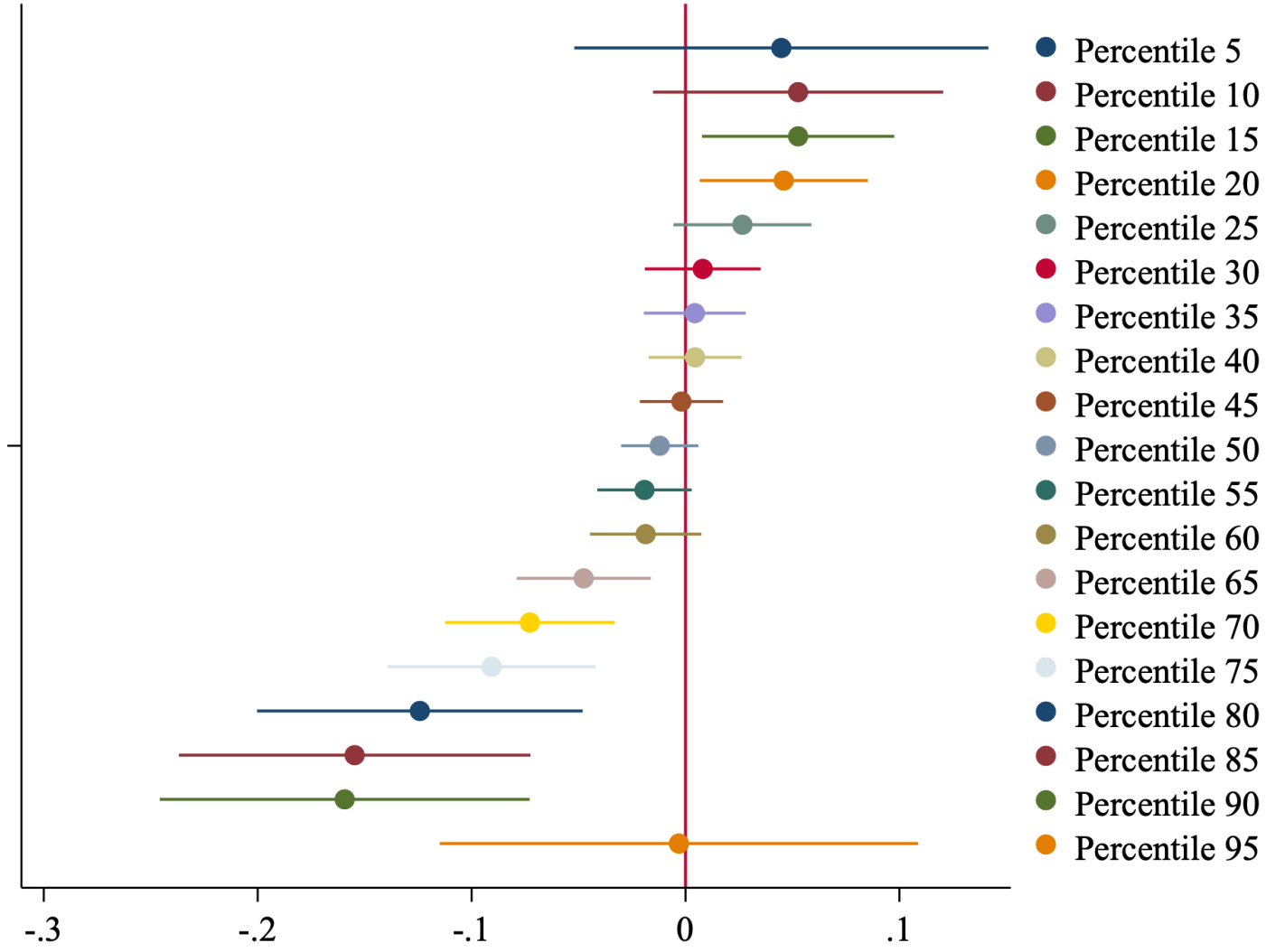
Notes: Plot comprising 9,116 responses over waves 2-10, of 3,052 firms that have at least one valid previous wave response. The forecasting error (arc-percentage) is calculated using 2*(Forecasted Sales Next 3 Months-Actual Sales Next 3 Months)/(Forecasted Sales Next 3 Months + Actual Sales Next 3 Months), 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.

Figure 4: Subjective Forecast Distributions are Over-Precise



Notes: Plot showing the percentile at which actual outcomes fall on the subjective elicited forecast distribution in periods 5-10 for 9,598 responses and 4,256 firms. Subjective forecast distribution is calculated using 3- and 5- bins of forecasted sales and probabilities, using the Altig et al. (2020) algorithm, as described in Section 3.2 and the Data Appendix.

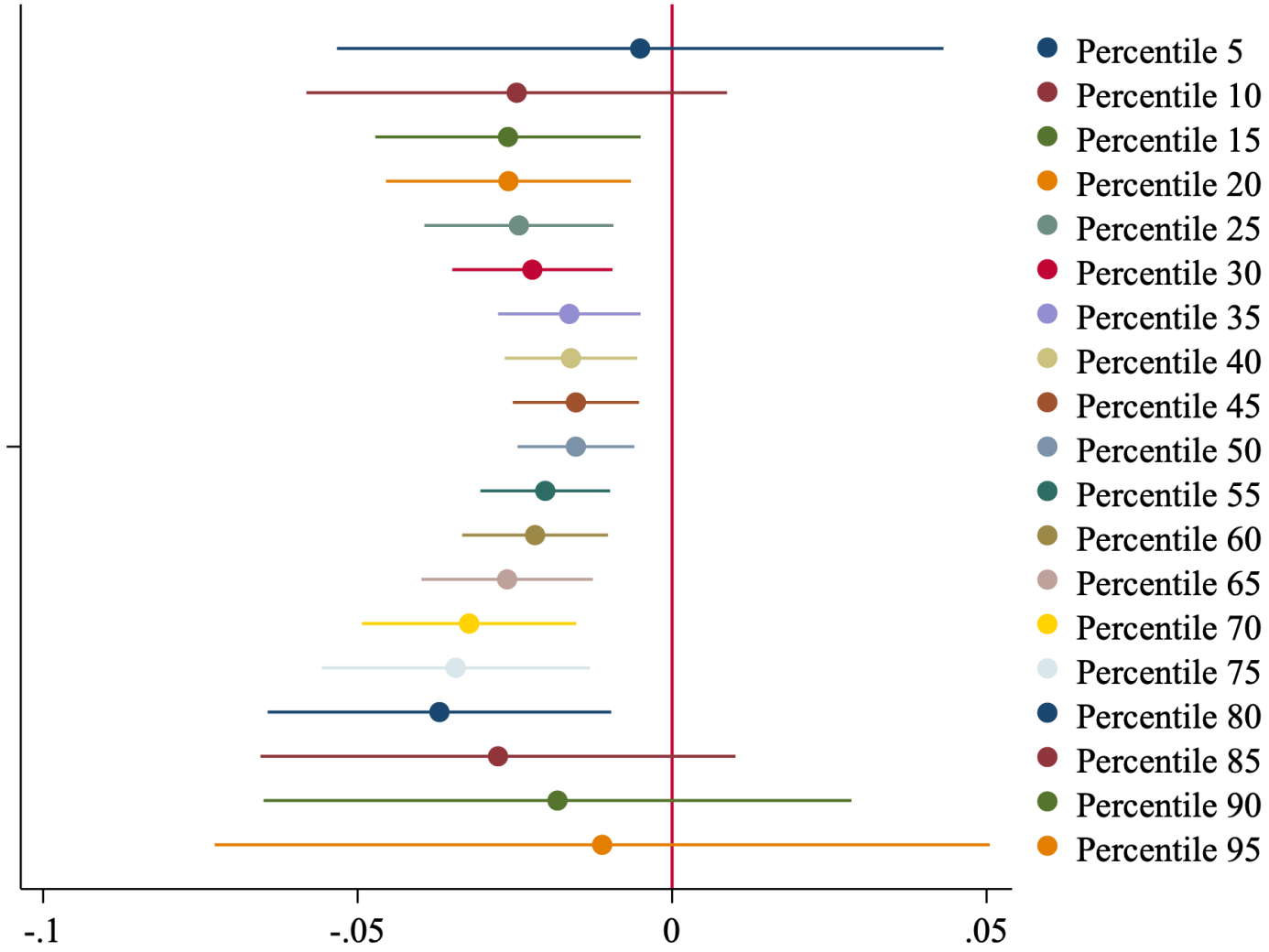
Figure 5: Quantiles of Dashboard Impact on Forecast Error



Coefficient (point) and 95% confidence interval (bar) by quantile

Notes: Quantile regressions showing 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.

Figure 6: Quantiles of Reward Impact on Forecast Error

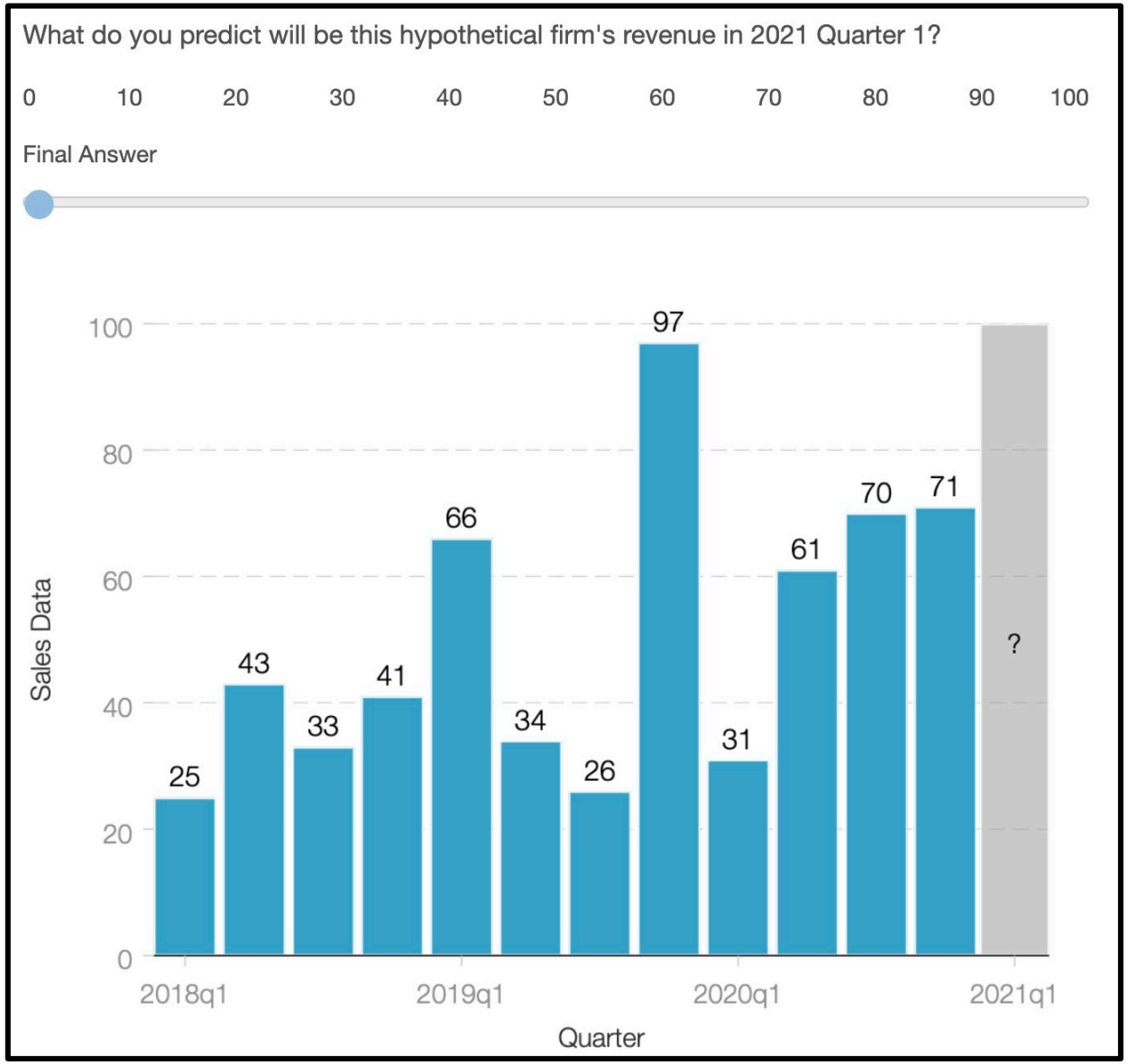


Coefficient (point) and 95% confidence interval (bar) by quantile

Notes: Quantile 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 5^{th}-95^{th} 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.

Figure 7: Forecasting training exercise

Screen 1



Screen 2

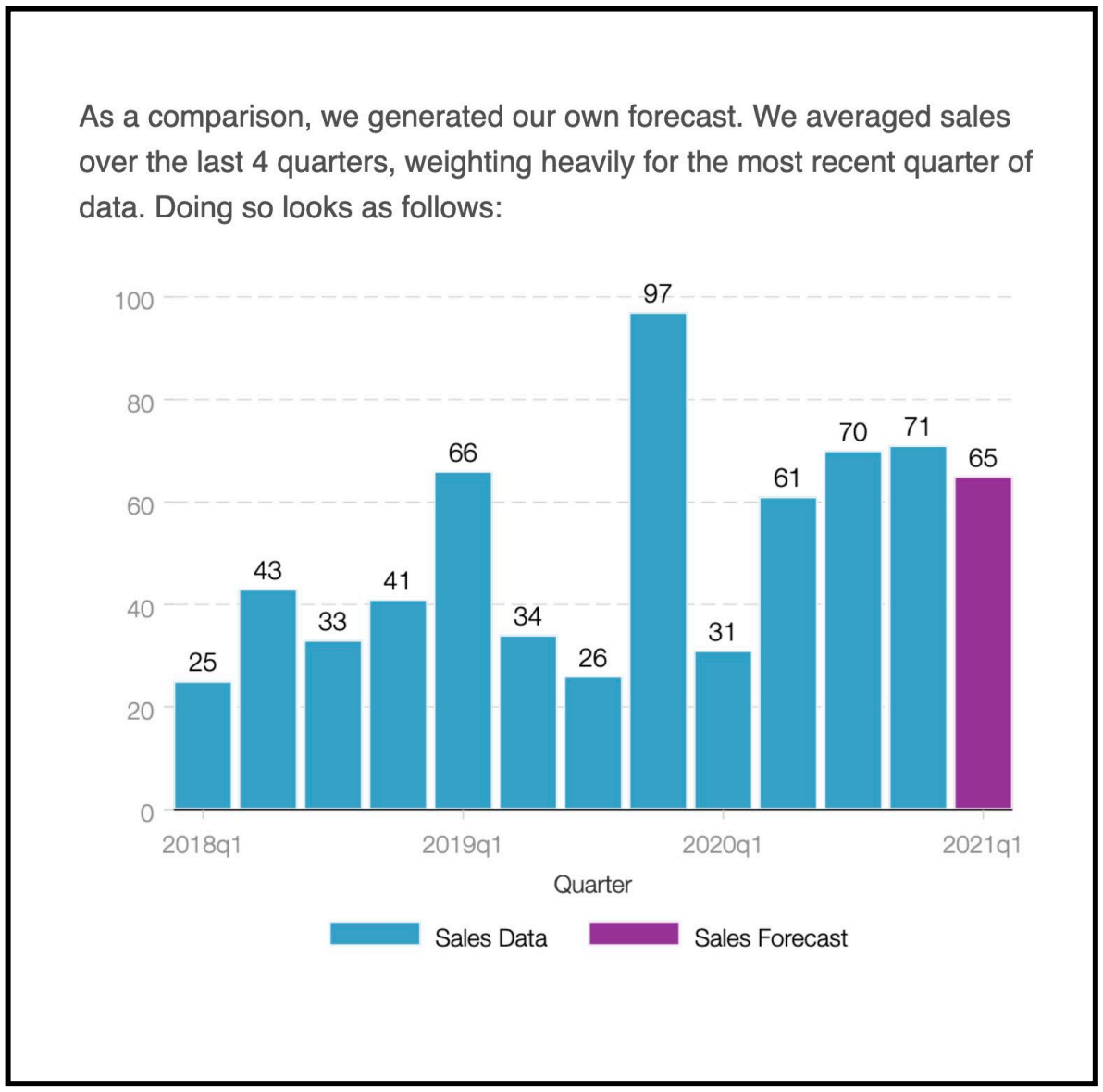
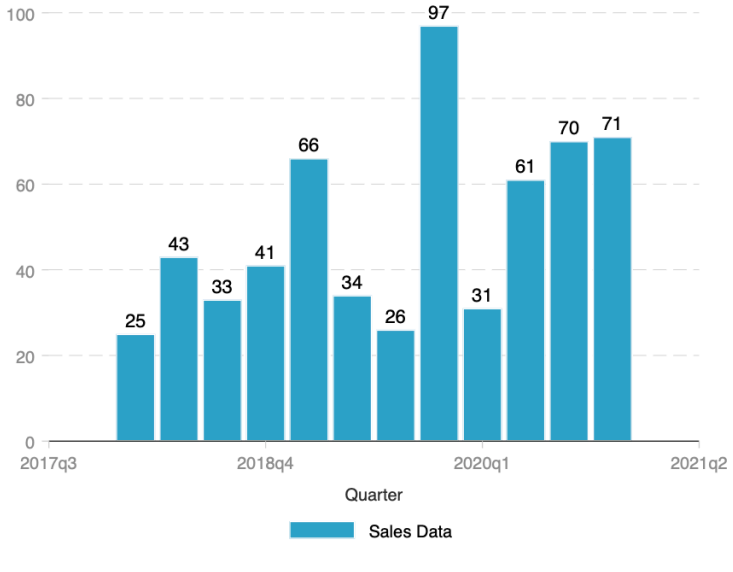
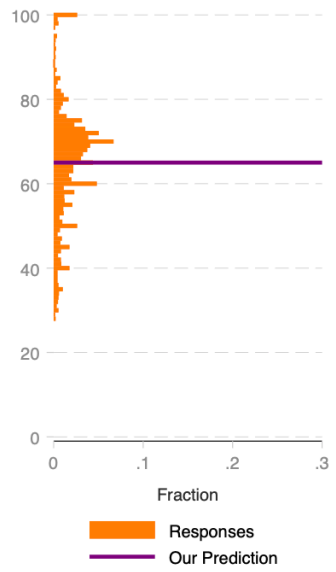
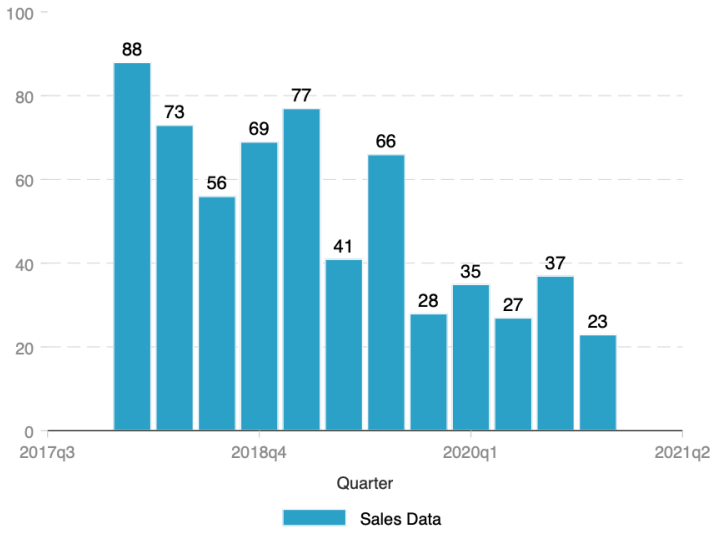
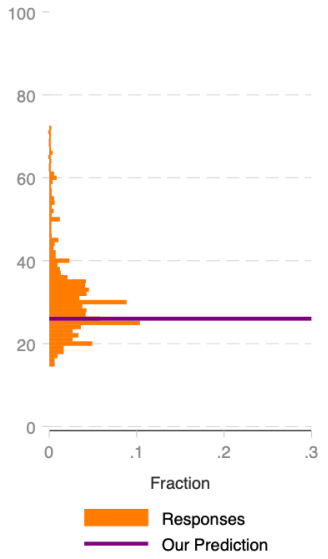


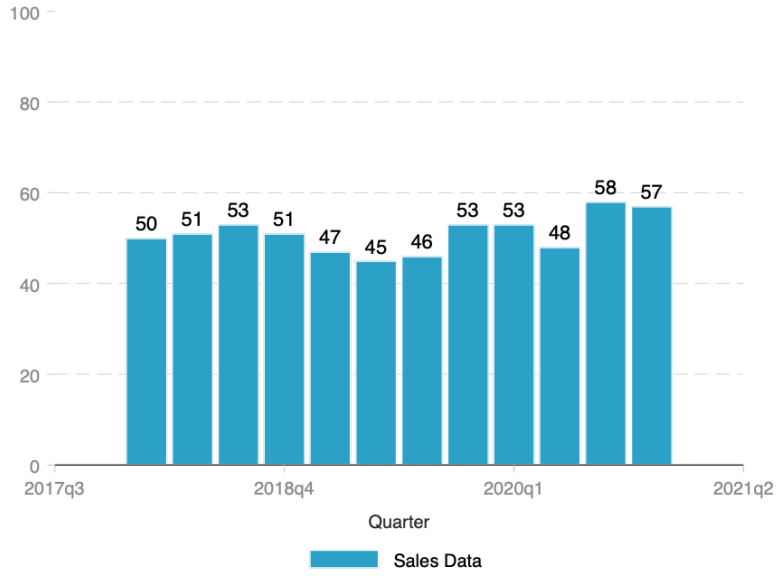
Figure 8: Example forecast training exercise results, and overall accuracy over the 10 exercises

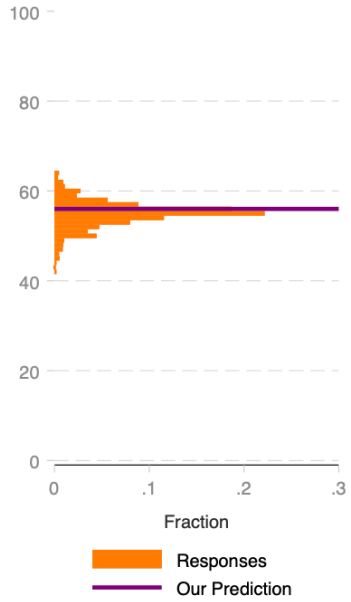


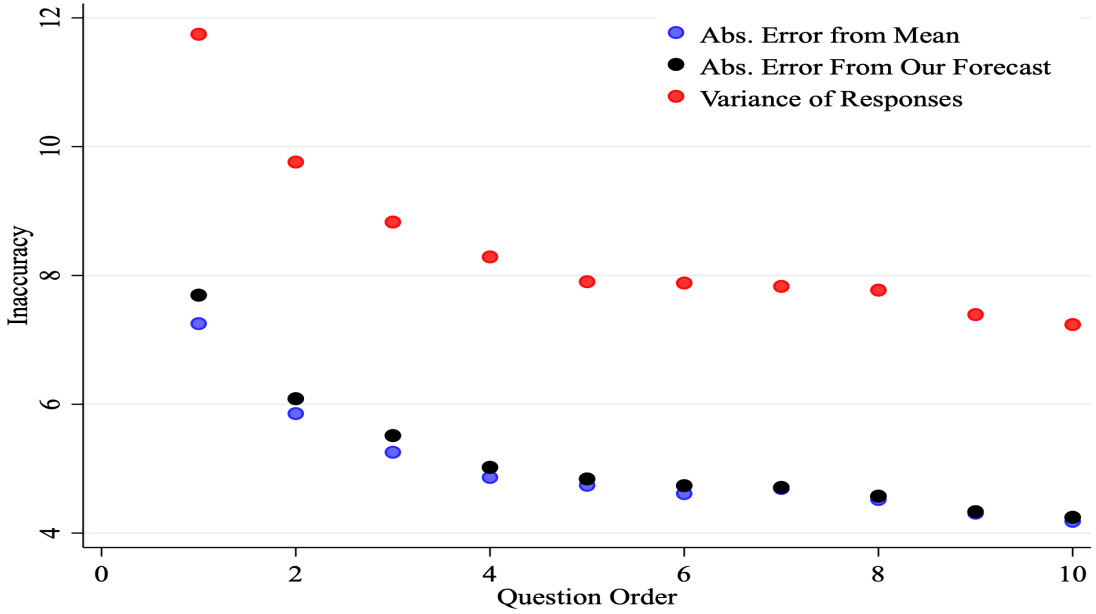






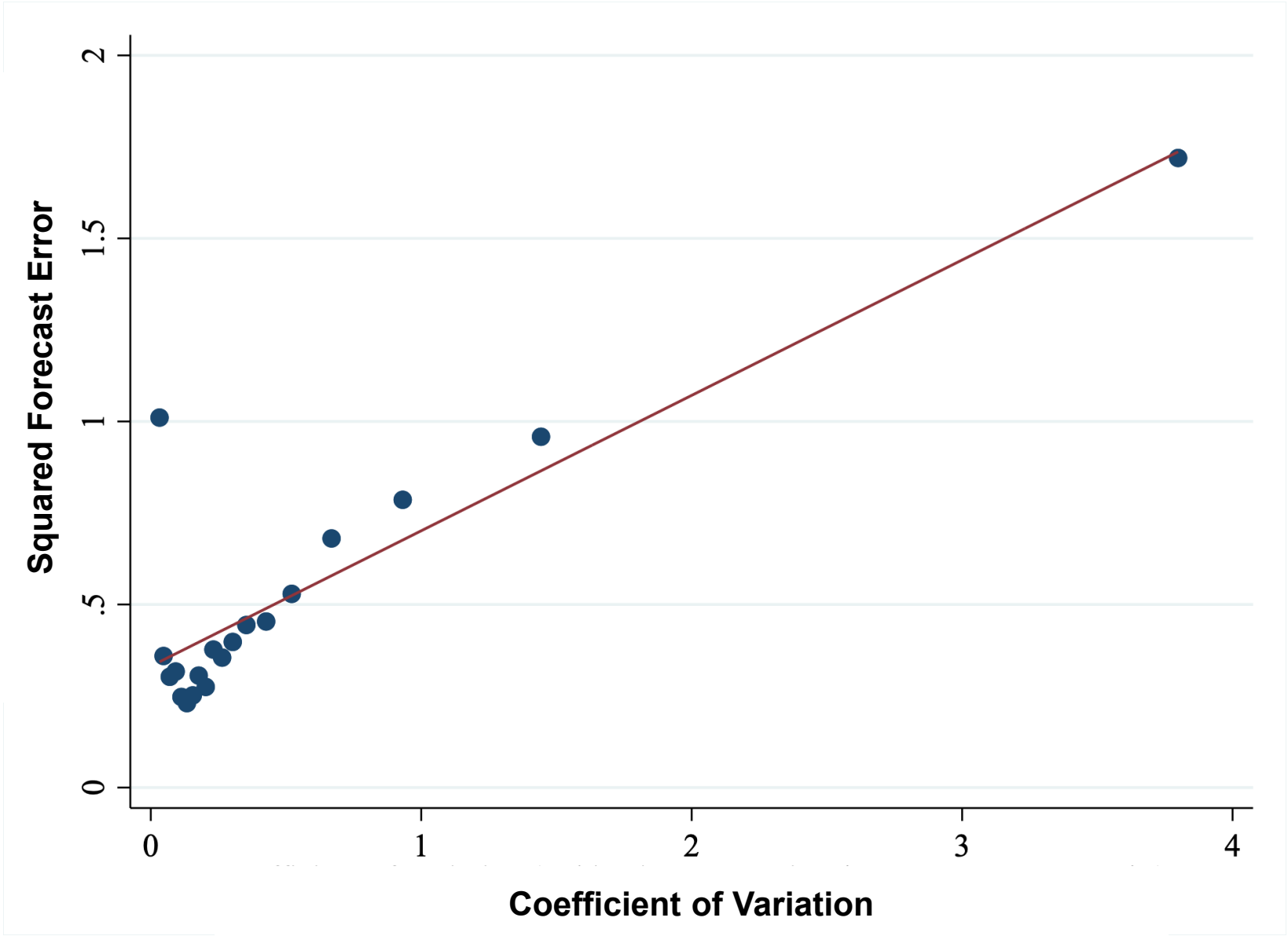






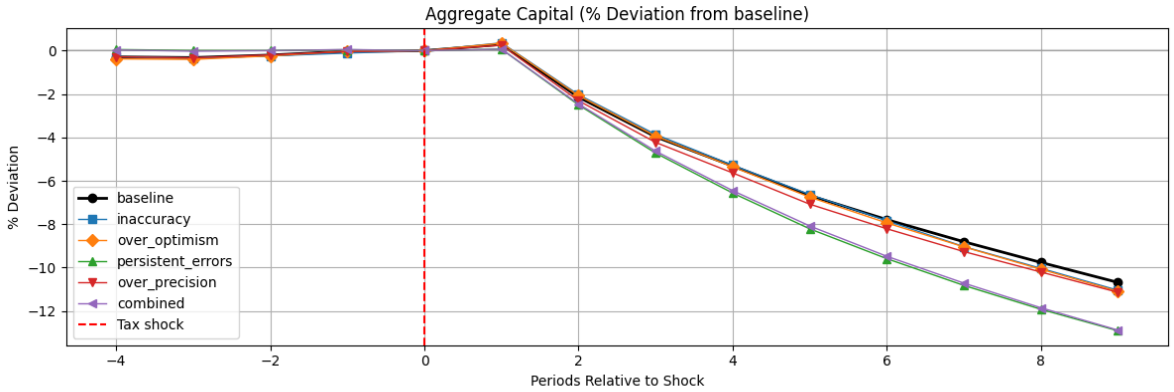
Notes. Three panels show training exercises and the distribution of individual forecasts. Final panel shows results over the 10 questions for the absolute error from mean, absolute error from our forecast and variance of responses, plotted against the randomized question order (N=27,412).

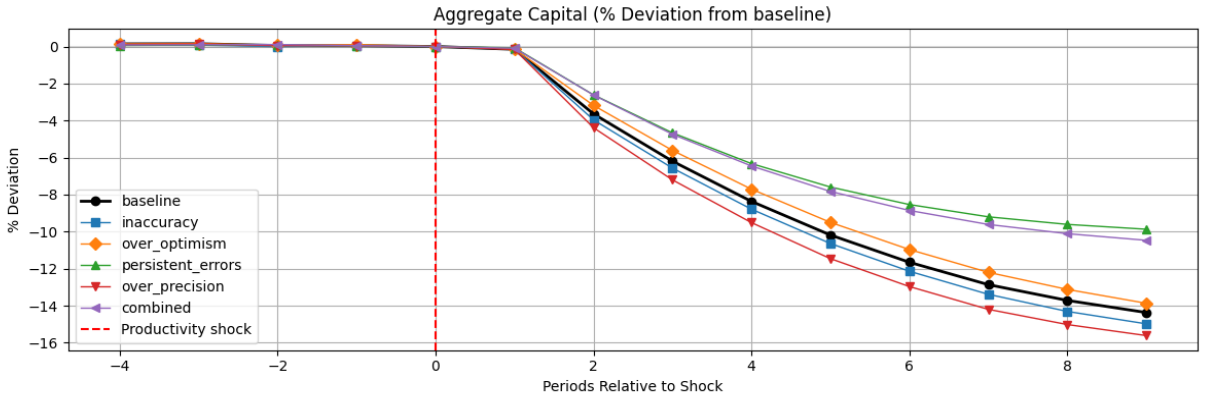
Figure 9: Scenarios subjective uncertainty vs forecast error

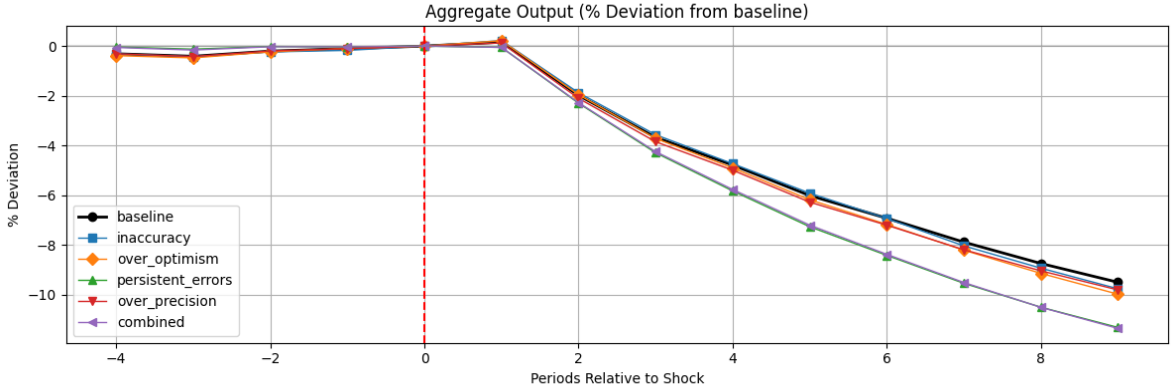


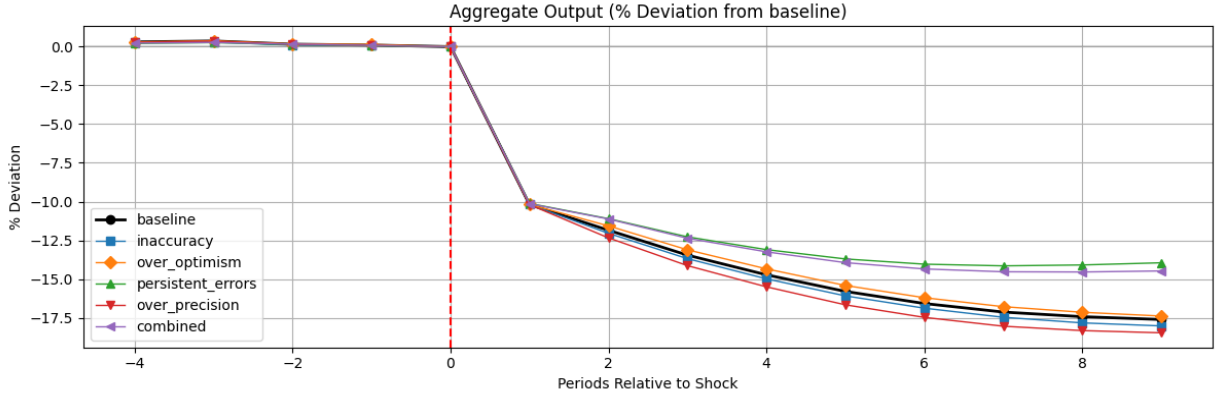
Notes. Binscatter (N=25) between squared 3-month forecasting error and the coefficient of variation, calculated from the subjective distribution from the 3-point or 5-point outcome probability forecast. The coefficient of variation is the subjective uncertainty measure, i.e. the standard deviation of 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).

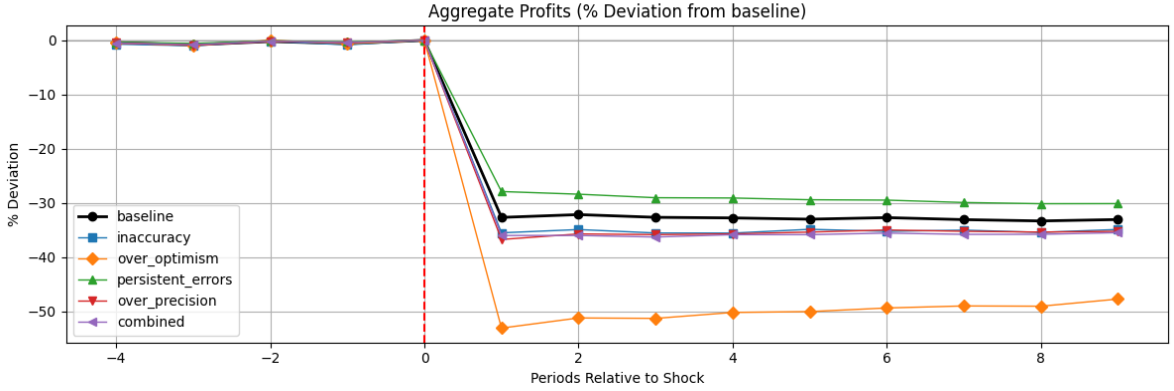
Figure 10: Impact of a tax rise (left) and productivity shock (right) by forecasting bias

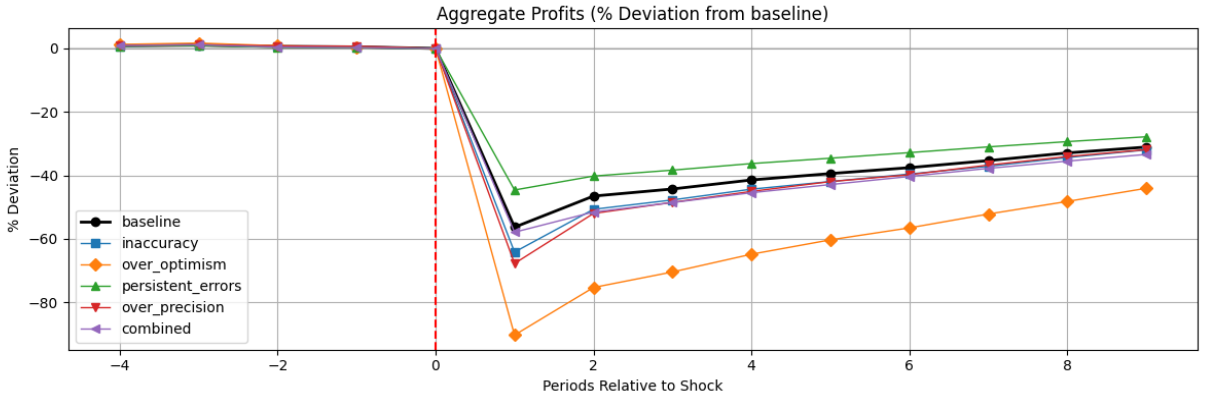








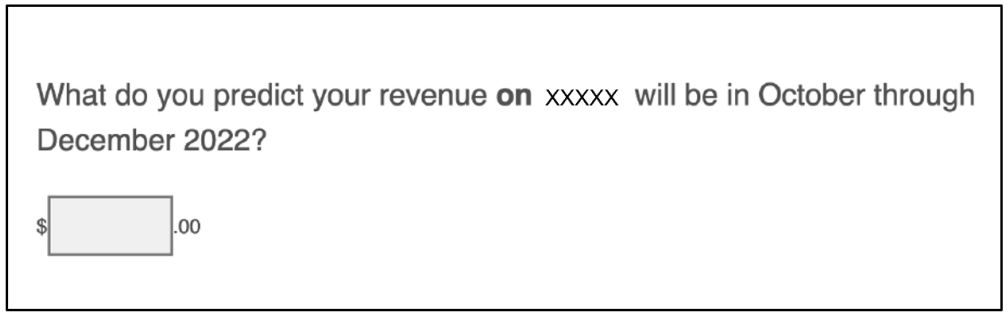


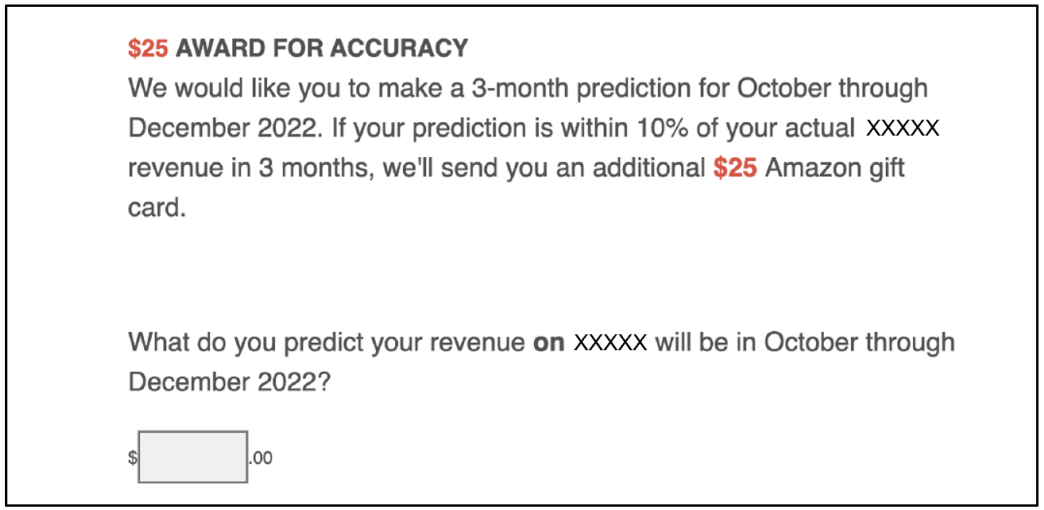


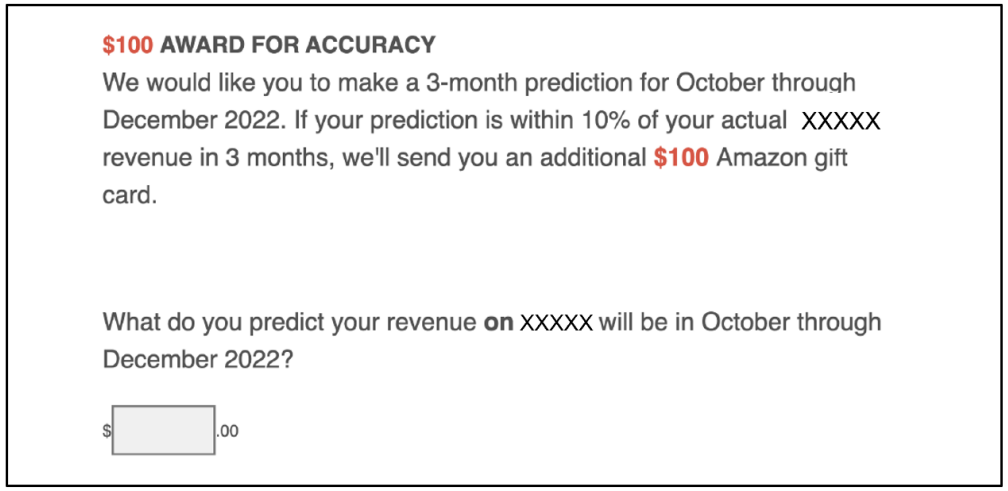
Notes. Impact of a 20% to 25% tax change (left) and 10% productivity shock (right) on aggregate capital, output, and profits, normalized to period t=0. Results from simulating 5,000 firms responses according to model described in section 5.

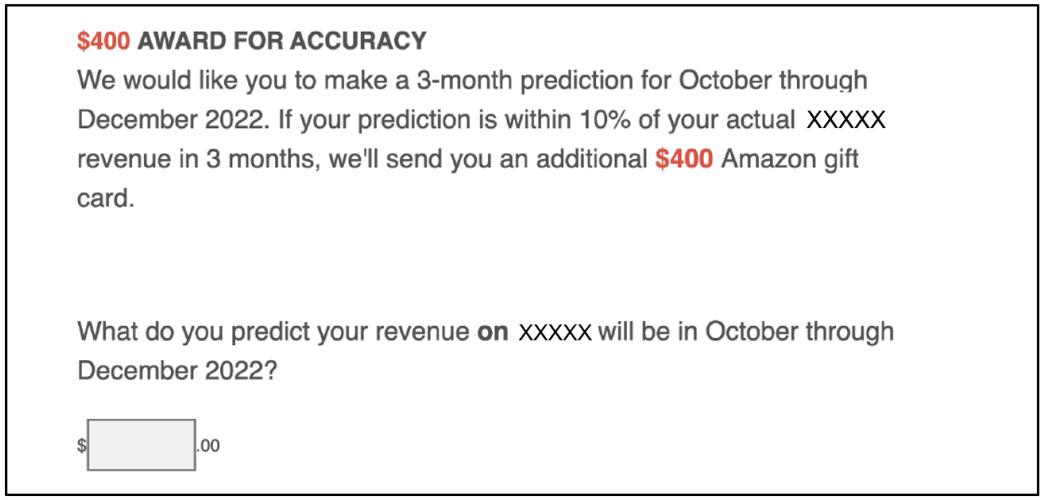
Appendix Figure A1: Forecast payment randomization from $0 to $400, four examples

In randomized forecast 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)



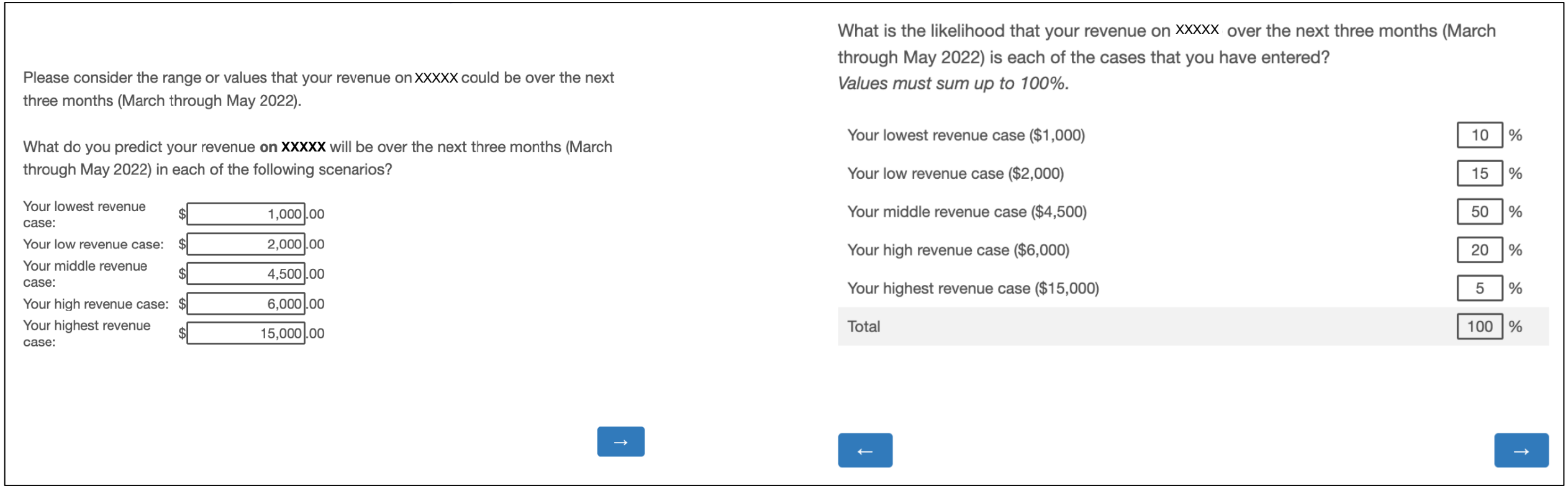






Notes: The name of the Fintech has been replaced with XXXXX in these screenshots. Rewards were varied with incentives from $0-$400.

Appendix Figure A2: Subjective sales forecast distribution question



Please consider the range or values that your revenue on XXXXX could be over the next three months (March through May 2022).

What do you predict your revenue **on XXXXX** will be over the next three months (March through May 2022) in each of the following scenarios?

Your lowest revenue case:	$ 1,000 .00
Your low revenue case:	$ 2,000 .00
Your middle revenue case:	$ 4,500 .00
Your high revenue case:	$ 6,000 .00
Your highest revenue case:	$ 15,000 .00

→

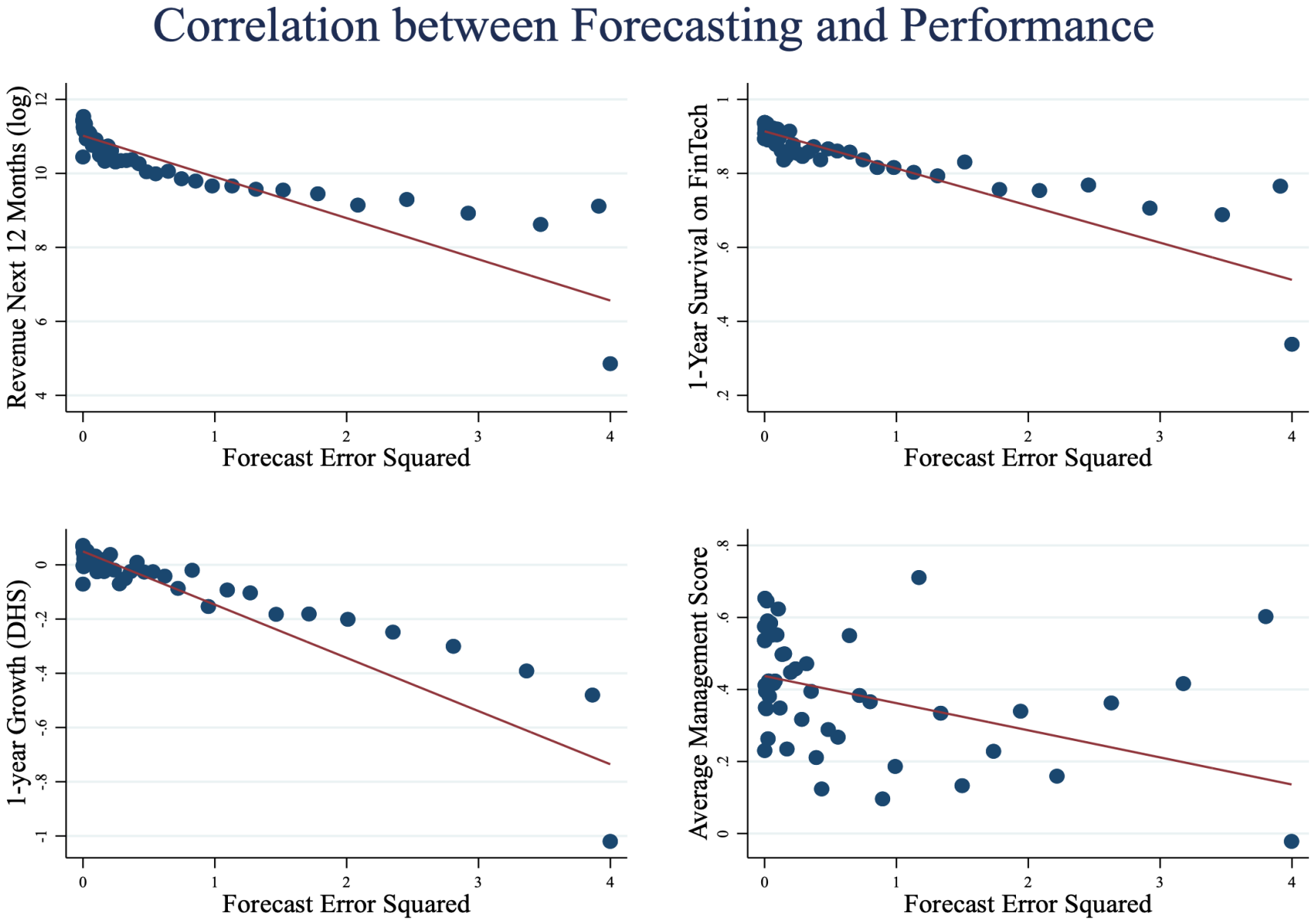
What is the likelihood that your revenue on XXXXX over the next three months (March through May 2022) is each of the cases that you have entered?
Values must sum up to 100%.

Your lowest revenue case ($1,000)	10 %
Your low revenue case ($2,000)	15 %
Your middle revenue case ($4,500)	50 %
Your high revenue case ($6,000)	20 %
Your highest revenue case ($15,000)	5 %
Total	100 %

← →

Notes: The name of the Fintech has been replaced with XXXXX in these screenshots. Shown for a hypothetical firm.

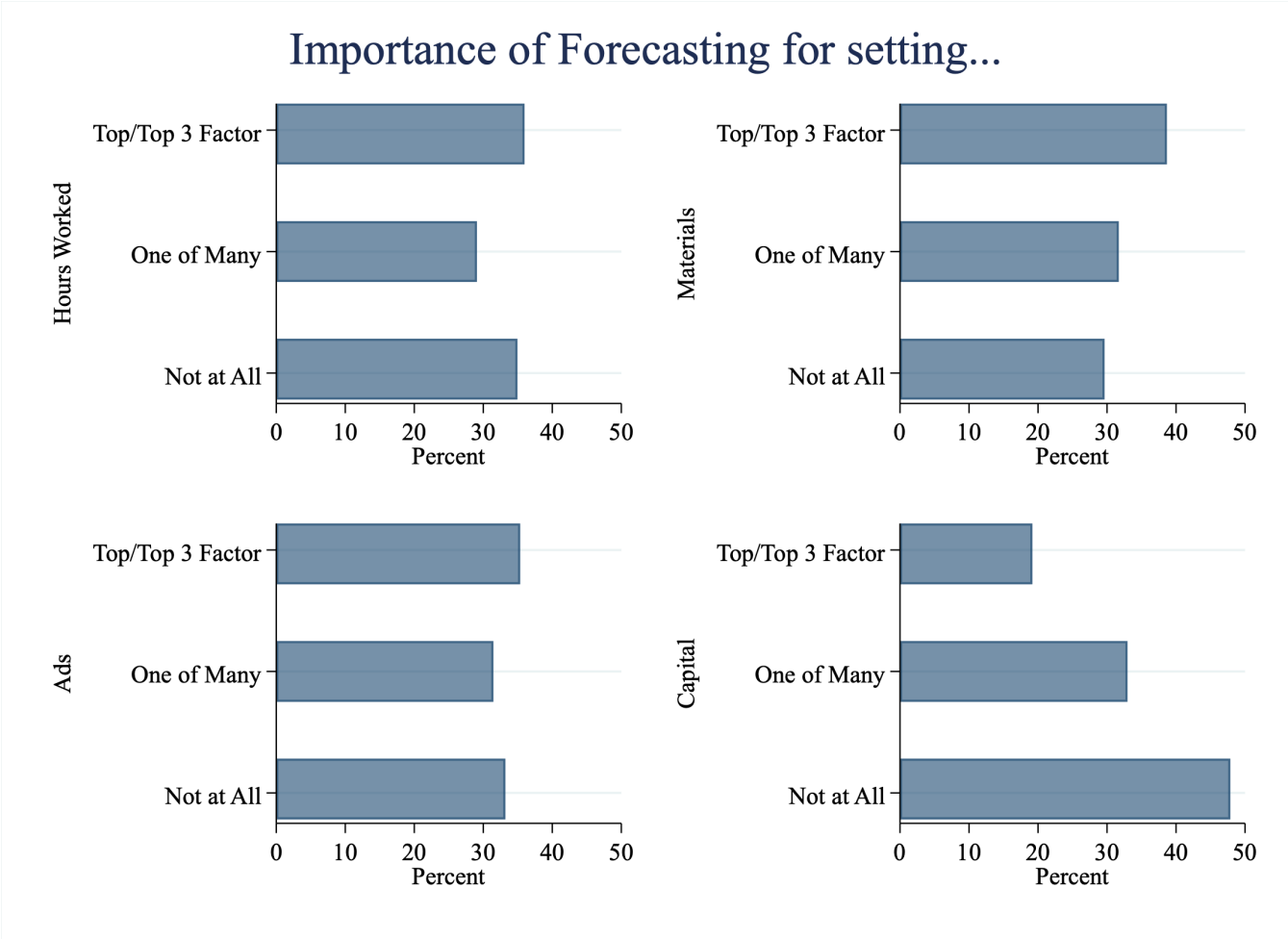
Appendix Figure A3: Performance and Forecasting



Notes: Panel showing binscatters of the correlations between revenue of the firm in the next 12 months (log), 1-year survival (any positive income) on FinTech, growth level (measured by 2*(Forecasted Sales Next 12 Months-Actual Sales Next 12 Months)/(Forecasted Sales Next 12 Months + Actual Sales Next 12 Months)) and average management score with the squared 3-month forecast error (arc-percentage). Results shown for the full sample when the full necessary data were available. The linear unconditional relationship between these measures gives t-statistics of -76.83, -45.28, -49.69, and -3.63.

Appendix Figure A4: Importance of Forecasting

Qu: “How important is sales forecasting for deciding [hours worked]”



Notes: Results for wave 8, over 1,262 responses of individuals who evaluated the usefulness of forecasting for their business in terms of setting hours worked, materials, advertising and capital.

Appendix Figure A5: A (fictional) example of the dashboard

