Rationalizing Entrepreneurs' Forecasts*

Nicholas Bloom † Mihai Codreanu ‡ Robert Fletcher §

January 10, 2023

Abstract

We analyze, benchmark, and run randomized controlled trials on a panel of 7,463 U.S. entrepreneurs making incentivized sales forecasts. We assess accuracy using a novel administrative dataset obtained in collaboration with a leading US payment processing firm. At baseline, only 13% of entrepreneurs can forecast their firm's sales in the next three months within 10% of the realized value, with 7.3% of the mean squared error attributable to bias and the remaining 92.7% attributable to noise. Our first intervention rewards entrepreneurs up to $400 for accurate forecasts, our second requires respondents to review historical sales data, and our third provides forecasting training. Increased reward payments significantly reduce bias but have no effect on noise, despite inducing entrepreneurs to spend more time answering. The historical sales data intervention has no effect on bias but significantly reduces noise. Since bias is only a minor part of forecasting errors, reward payments have small effects on mean squared error, while the historical data intervention reduces it by 12.4%. The training intervention has negligible effects on bias, noise, and ultimately mean squared error. Our results suggest that while offering financial incentives make forecasts more realistic, firms may not fully realise the benefits of having easy access to past performance data.

Keywords: management, uncertainty, sales, forecasting, firm performance
JEL Codes: L2, M2, O32, O33

*We thank the Kauffman Foundation for financial support, Eilin Francis, Nate Hilger, and Alex Novet for their diligent work in helping to establish the survey, Ethan Yeh and Ana-Maria Mocanu for extensive larger project support and advice. We are grateful to seminar participants at the NBER Summer Institute, Columbia Business School, Stanford University, Chinese University of Hong Kong, Stripe Data Science, and the Census Bureau. The analysis presented in the paper are authors' only and do not necessarily reflect the views of Stripe, Inc.

†Stanford University (Department of Economics) and NBER, email: nbloom@stanford.edu
‡Stanford University (Department of Economics), email: mihaic@stanford.edu
§Stanford University (Department of Economics) and Instacart, email: rsfletch@stanford.edu

1 Introduction

Managerial expectations are both a key determinant of firm investment and production decisions and a key input into the design of fiscal and monetary policy. However, empirical research on the expectations of firm managers has been limited. By and large, economists assumed firm managers had rational expectations and so sidestepped the issue until recently (Gennaioli, Ma, and Shleifer 2016).

With the advent of the literature showing the heterogeneity of management attributes and practices and their important impact on productivity,[1] a surge in studies have sought to empirically characterize manager forecasts directly (e.g., Bachmann et al. (2020), Altig et al. (2019)). These studies have mostly relied on primarily non-incentivized direct elicitation of forecasts from business managers analyzed against mostly self-reported or long-run company financial results, with the goal of identifying psychological biases in the forecasts that disprove rationality (e.g. managerial overconfidence).[2]

We build on this work by using panel survey evidence, linked to administrative data, in an incentivized prediction setting, to examine the types of errors that entrepreneurs make and assess which categories of interventions are likely to help improve their performance. This survey has been carried out in partnership with a leading U.S. financial tech company, which we refer to as TechCo, whose detailed administrative sales data allows us to verify the accuracy of sales forecasts over the short-term (3 months) and longer-term (12 months). To our knowledge, it is among the first to explore and experimentally evaluate types of interventions that aid entrepreneurs in making forecasts.

Our first contribution is a characterization of entrepreneur forecasts. Despite the apparent importance of sales forecasting accuracy (around 90% of sample reporting they are a top/top 3 factor for deciding at least one of hours worked, material, capital or advertis-

1. The literature in this subfield is rising fast and ranges from older studies analyzing the relationship between manager style and firm policies (Bertrand and Schoar 2003), to newer literature on e.g. the longer term impact on firm performance (Giorcelli 2019; Bloom et al. 2019), employee attrition (Hoffman and Tadelis 2021), as well as various randomized controlled trials seeking to improve firm productivity via interventions at the managerial level (Gosnell, List, and Metcalfe 2020), including in developing countries (Anderson and McKenzie 2022; Adhvaryu, Kala, and Nyshadham 2022; Bruhn, Karlan, and Schoar 2018)

2. There are three types of overconfidence biases routinely discussed in the behavioral literature. These are (Moore and Healy (2008), Santos-Pinto and Rosa (2020)): overestimation (of one's absolute skills), overplacement (of one's relative skills), and overprecision (overestimation of the estimates/knowledge precision).

ing), entrepreneurs are remarkably inaccurate.[3] We reward entrepreneurs for sales forecasts within 10% of the realized value for the short-run predictions, and we still find that only 13% entrepreneurs successfully predict within this window, at baseline.[4] To benchmark these numbers, a random walk model inputting the sales from the previous three months as the prediction would achieve an accuracy of 15% at baseline.

Following Kahneman, Sibony, and Sunstein (2021), we additionally analyze forecasting errors using a simple bias-noise (bias-variance) decomposition.[5] Despite the focus in the literature on overoptimism of managers and entrepreneurs, we find in our data, at baseline that noise drives 92.7% of forecast mean squared error (MSE) and (upward) bias only contributes 7.3%. The percentage of MSE explained by noise rises to up to 99% in some of the waves affected by the Covid-19 shock. These results are in line with the work of Kahneman, Sibony, and Sunstein (2021), which suggests that when social scientists study forecasts and judgements, they routinely place too much emphasis on bias despite noise causing the lion's share of aggregate error.

Given these characterizations that forecasting quality is poor and almost entirely because of noise, we turn to our central research question: what types of treatments best suit the types of errors entrepreneurs make? We propose classifying interventions by their mechanism: attention, data-quality, and skill. Attention refers to treatments that elicit greater amounts of time and effort from forecasters who may be under-investing in forecasting. Data-quality interventions improve the quality of data and information used as a basis for the forecast. Skill interventions seek to improve the forecaster's ability to better leverage information and

3. The importance of accurate forecasts was highlighted, from a theoretical point of view, by research as old as Tobin's Q-theory of investment (Hayashi 1982). Relatedly, Bloom et al. (2019) finds that more productive and better managed firms have improved forecast accuracy in the U.S., while Massenot and Pettinicchi (2018) and Bachmann and Elstner (2015) use German manufacturing data and find that larger and older firms have smaller forecasting errors. Despite the theoretical consensus of the importance of accurate forecasts, some empirical papers associate managerial overconfidence with benefits such as firm-level innovation in Galasso and Simcoe 2011 and Hirshleifer, Low, and Teoh 2012, increased corporate investment (Ben-David, Graham, and Harvey 2013)Others highlight costs such as overpayment for target companies as and curtailment of investment that requires external funds (Malmendier and Tate 2005, 2008), sorting into more performance-driven contracts (Larkin and Leider 2012), and underinvestment in information production (Goel and Thakor 2008). Notwithstanding measurement and selection issues, there are a few papers highlighting mechanisms through which these findings could be explained, for example due to differential risk attitudes (e.g. Bruhin, Santos-Pinto, and Staubli (2018)).

4. The baseline success figure refers to the first wave of the survey, completely preceding the Covid-19 shock, but results are consistent along the later waves, with success rates generally below 20%.

5. This refers to the decomposition of mean squared error into the sum of bias squared and variance.

data through the use of superior methods or heuristics. For each of the classes, we run an experiment. In order, we offer rewards for accuracy, guide respondents through to review their historical data, and lead them in a forecasting training.

The first class of interventions (i.e. attention-driven) encompasses the distinct ideas of increasing effort devoted to the forecast and willingness to adhere to realistic responses. Entrepreneurs may feel that forecasts are not important for them and their business, or that realism is not of great value in a survey setting, which could explain their poor performance. To incentivize greater effort, we randomize the amount that individuals receive for a correct forecast, with amounts varying between \$0 and \$400, instead of \$25 at baseline. Offering higher rewards may induce individuals to exert more time and attention in their forecasts thus reducing overall forecasting errors. Second, paying individuals even small amounts can induce respondents to offer more realistic responses to questions that they have an emotional tie to, as in the work of Bullock et al. (2015).

We find that when respondents are offered higher amounts, they spend significantly more time on the predictions. They are also more realistic in their forecasts, as we see a significant decrease in the upward bias coming from manager overoptimism (by around a third). We do not see a reduction in the much larger issue of noise of their forecasts, with the average squared forecast error remaining at a similar high level.

The second class of interventions (i.e. data-driven) addresses directly that entrepreneurs may not be able to forecast accurately given a lack of information on which to form their forecasts. To test this, we randomly assigned entrepreneurs to review their historical data immediately before being asked to give a forecast. This was done using the financial dashboards that TechCo provides to all of the users in our sample. These dashboards are on the front page of their account and clearly display financial metrics, most notably past sales. We document that prior to the intervention, only 33% of entrepreneurs can accurately report their past revenue, which suggests that they are not using their dashboards.

The results of the data intervention showed substantial reductions in mean-squared error by 12.4%, primarily via reducing noise. The intervention was not effective in terms of reducing mean overconfidence. These results suggests that a potentially easy and effective way of decreasing noise in managerial forecasts is increasing the availability of data, and encouraging its usage.

Last of all, we tested skill interventions with a training module. A random subset of respondents were guided through making forecasts on hypothetical sales data. This training was simplistic and was focused on the basic aggregation of historical data so as to address first-order trends in the data (at a theoretical level, AR(k) with $k \leq 4$ type of models). The results of this intervention were non-significant, with no noticeable change in either bias or noise. We interpret these results as reflective of the difficulty of making accurate forecasts for the businesses, whose sales are extremely volatile from quarter to quarter, especially in the context of insufficient previous data knowledge. Beyond the basic use of data in the forecasts, gains in accuracy from improved forecasting methods are hard to achieve.

In sum, entrepreneurs perform poorly at forecasting, even when given large incentives for accuracy and despite having effective and easily accessible data tools at their disposal. This begs the question of why these entrepreneurs do not leverage tools such as the dashboard which are already available to them.

We find it unlikely that it is because forecasting is not valuable for entrepreneurs, or that they are not willing to engage with our incentivized predictions. For one, we have just shown that entrepreneurs are willing to spend more time when they have higher rewards but their predictions in terms of noise do not improve markedly. Combined with the significant impact of the dashboard treatment, this suggest that entrepreneurs may not realize that dashboard usage could help them get better predictions. Beyond this, we also find that the ability to correctly forecast sales is highly positively correlated with observed firm performance, even when controlling for other management characteristics scores.[6][7] This result also holds when using forecasting ability as measured within our training module, an objective, own firm characteristics independent metric. Taken together, these results suggest that forecasting accuracy plays a large role in firm performance and that firms may be undervaluing the use of readily-available data when making their forecasts.

Last of all, we also show that entrepreneurs are significantly upper biased in their perceived forecast accuracy. This is true at a general sense, with respondents giving overprecise

6. This result is similar to the results found by Tanaka et al. (2019), with the notable difference that they concentrate on the connection between macro forecast accuracy and firm performance, while the managers in our sample are predicting their overall own firm performance. They likewise find a strong relationship between forecast accuracy and firm performance.

7. Other studies have attempted to causally estimate the impact of forecast accuracy on firm performance using models with adjustment costs following inaccurate forecasts (e.g., Asker, Collard-Wexler, and De Loecker (2014), David and Venkateswaran (2019), Ma, Sraer, and Thesmar (2020)). The results in this literature are more mixed. In subsequent work we will test if the effectiveness of the interventions also influences future firm performance.

forecast ranges and vastly overestimating their odds of giving a winning forecast, and also in a relative sense – self-reported forecast ability is negatively correlated with actual forecasting ability. For example, the actual sales in the next 3-months are outside of the reported worst-best case scenario interval a striking 39% of the time. When asked for 90% confidence interval, less than 60% of realizations fall within this window. Very likely, this overconfidence in their forecasting ability causes them to undervalue the adoption of more rigorous data-driven judgment.

Our paper and its findings relate to a number of strands in the literature, the two most important being manager forecasting and the evaluation of interventions to improve forecasting performance. Papers that have studied manager forecasting fall into a few main categories. First, there are those papers which ask managers to forecast macroeconomic data like inflation (e.g., Malmendier and Tate (2015), Malmendier and Nagel (2011), Coibion, Gorodnichenko, and Ropele (2020), Coibion and Gorodnichenko (2012)). Second, there are a number of studies that analyze managers expectations' for their own firms (Barrero (2021), Hebert (2021), Ben-David, Graham, and Harvey (2013), Altig et al. (2019)). To this literature, we contribute experimental evidence on improving forecasting errors and an analysis of why certain classes of interventions are most likely to succeed, using new administrative data.

The literature on the evaluation of forecasting methods and interventions is much shorter, given the logistic difficulty in running large scale forecasting experiments. Our paper is most similar to the work in Mellers et al. (2014) and Satopää et al. (2021), which analyze the effect of various interventions on global event forecasting performance. Satopää et al. (2021) similarly finds that noise plays a major role in global events forecasting errors and that the most successful interventions are those that are able to tackle noise directly.

In the rest of this paper, we expand on each of the above points. Section 2 provides an overview of the survey and TechCo's administrative data used in the analysis. Section 3 looks at forecasting performance as it stands. Section 4 analyzes the randomized controlled trials of our interventions with results on how they affect forecasting performance. Section 5 considers the aggregate value of accurate forecasting for entrepreneurs. Section 6 and 7 concludes with a discussion of our results, further avenues to pursue and a presentation of our forthcoming plans for the next waves of the survey.

2 Data

2.1 Survey Design

The Study of Internet Entrepreneurship Survey is an opt-in panel survey of business founders in partnership with a large payments technology company in the United States, also producing leading payment processing software for e-commerce websites and mobile applications, henceforth referred to as TechCo.[8] The sample was constructed from the universe of businesses using TechCo's online payment services. To be eligible for the survey, businesses had to have had at least ten transactions on TechCo. To limit the inclusion of businesses that had already closed, they also had to have had a transaction in the 90 days prior to when they were sampled. Businesses had to be for-profits, and the emails that TechCo had listed for them had to be non-generic.[9]

Our surveys were targeted at business founders. If the founder was not available or was no longer affiliated with the business, then we accepted the responses of someone who was intimately familiar with the financials of the company and the TechCo account itself. In 92% of responses, we were able to get a response from the founder themselves.

So far, there have been a total of 8 rounds of the survey, with 2 more rounds planned for 2022 and 2023, as covered in Table 1. Each round of the survey consists of a core set of questions on the finances of the business, labor use, and forecasting. On top of this, there is an updating series of modules asking about one-off topics such as management, COVID-19 impacts, personality traits, and other shorter points of interest. The waves eliciting the Covid-19 effects (waves 4 and 5) are presented thoroughly in Bloom, Fletcher, and Yeh (2021).

The eligible firms were divided into three strata: funded, small non-funded and large non-funded. Funded firms were those known to have venture capital backing. Non-funded firms were then split into small and large based on the amount of revenue they had on TechCo in the prior year. Firms with below $10,000 in revenue the previous year were labeled small, while firms above $10,000 were labeled large. Our initial sample size was

8. To facilitate this research, TechCo allowed the authors to communicate with its customers to request their survey participation and limited, anonymized access to data from users that granted permission and opted-in to the study.

9. In practical terms, they could not consist of phrases such as "info@", "admin@", or "contact@"

Table 1: Survey Rounds Overview

Round	Dates	Responses	Extra Module	Interventions
1	Jan-Apr 2019	3,941	Baseline Characteristics	
2	May-Aug 2019	2,891	Management	
3	Oct 2019-Jan 2020	3,185	Personality	
4	Apr-May 2020	2,446	COVID-19 Part I	
5	Sep-Oct 2020	2,409	COVID-19 Part II	Dashboard + Reward
6	Jan-Apr 2021	1,883	Management	Dashboard
7	Sep-Nov 2021	3,100	Forecasting	Dashboard + Reward + Forecast Training
8	Apr-Aug 2022	1,938	Forecasting Importance	Forecast Training II
9	Sep-Dec 2022	TBC	TBC	TBC
10	Jan-Mar 2023	TBC	End of survey	

Notes: Data for firms comes from 7,463 surveyed entrepreneurs in the Stanford Study of Internet Entrepreneurship.

made up of a third funded, a third small, and a third large firms.[10] More information about the basic characteristics of the respondents and the participating firms can be found in the next subsection, as well as in Table 2.

We contacted a total of 46,400 firms. Firms were contacted with an invitation e-mail on their official TechCo account and three follow-ups spaced approximately a week apart. Respondents were given either $25 or $50 to respond to the first wave of the survey and then $25 for each subsequent wave. In addition, as will be discussed below, they were also given an additional award per survey wave for accuracy for their TechCo revenues forecasts for the next quarter if they came within 10% of their realized revenue (which was set as baseline for $25).[11]

Firms who did not respond were then contacted again in the following round of the panel and reinvited to participate with an invitation and two reminder emails. A total of 7,463 firms responded throughout, for a response rate of 16.1%.[12] We found that the most

10. Note final sample size contained a smaller share of funded firms since all of the eligible TechCo funded companies were invited to complete the survey by 2021.

11. The $25 amount was randomized as part of one of our interventions in rounds 5 and 7.

12. While this 16.1% response rate may seem low, it is high for firm surveys, especially during the pandemic. Prior COVID-19 firms surveys obtained response rates that were substantially lower, for example 0.017%

significant difference across firms was their size, with smaller firms more likely to respond (see Table A.1).[13] We contacted 18,000 businesses throughout the spring of 2019. Firms were then re-contacted in the summer of 2019. Those who had not completed the first round were re-invited to take the baseline survey, while those who had already completed the baseline survey were given the second-round survey. The third round of the survey took place at the end of 2019. Firms who had only completed the baseline survey were invited to complete the third round with the other firms, thus skipping the second round. We also refreshed our sample with an additional 4,400 businesses at this point, giving us a total number of 22,400 firm which were contacted.

A fourth round was then sent out during April and May 2020. This round coincided with the onset of the COVID-19 pandemic, and so included the questions on the impact of the crisis which form the basis of our COVID-19 analysis. A fifth round was then sent out during September through November 2020. This round followed the peak of the COVID-19 economic impact and so allows us to analyze retrospective data, as well as compare forecasted and actual impacts. As in previous rounds, we added an additional refresh sample of 4,000 firms for a full sample of 26,400 firms. In the sixth round of the survey, we covered the period between April to May 2021.We then sent out our latest wave with 3-month forecasts that were fully analyzed (seventh) covering September through November 2021. We augmented this wave with a much larger refresh sample of 20,000. Waves eight and nine are set to be completed in 2022 without a refresh sample. A full timeline is available in Table 1.

2.2 Founder and Business Characteristics

From the baseline survey, we collected a number of characteristics on the founder and their business. Table 2 shows some basic characteristics, while Figure A.2 compares them to businesses from the Annual Survey of Entrepreneurs (henceforth ASE), which is a nationally representative survey of all (rather than TechCo) businesses. We see the average entrepreneur in our survey is 39 years old, slightly below the 42 years of age of the average US entrepreneur (from the ASE), with 95% of TechCo business leaders more than 25 years old. We also see that 72% of firms are run by college graduates, reflecting the increasing importance of education for entrepreneurship in the new tech intensive economy. Finally, most of these

for Bartik et al. (2020) and 1.5% in Alekseev et al. (2020), while pre-pandemic US firm surveys typically obtained response rates between 10% to 30% (e.g., see Altig et al. (2019)).

13. We are currently checking the robustness of our results by employing inverse probability weighting for the characteristics that predict survey responses. Initial results confirm the findings in this draft.

firms are young, with 65% of them having been founded within the last 5 years, in contrast to all US firms which have an average age of 17 years. As expected, they have a very high percentage of their revenue earned online, just in excess of two thirds.

Table 2: Summary statistics at survey entry

	Sample size	Average	Median	Std. Dev.	Min	Max
Firm Characteristics						
Number Founders	6630	1.5	1	0.8	1	5
Number Employees	6630	10.4	2	203.4	1	16000
% Revenue Online	6630	67.5	90	37.6	0	100
% Revenue TechCo	6630	52.1	50	36.1	0	100
% Revenue International	4586	8.6	0	19.0	0	100
Revenue past 12 mo. ('000)	6630	403.8	80	795.2	2	3070
TechCo Revenue past 12 mo. ('000)	6630	151.7	24	337.6	0	1400
Firm Age	6579	5.9	4	6.0	0	81
Funded Flag	6630	0.20	0	0.40	0	1
Entrepreneur Characteristics						
Age	6630	39.2	37	10.7	16	100
Hours worked (per week)	6630	40.3	40	22.2	0	100
Earnings from firm past 12 mo.	6630	51.5	30	60.3	0	215
Number Businesses Owned	6630	1.5	1	0.8	1	5
Number Previous Businesses	6630	1.0	0	1.3	0	5
Has Other Job Flag	6630	0.3	0	0.4	0	1
Total sample size	7463					
Valid sample size	6630					

Notes: Data comes from 7,463 unique entrepreneur survey responses, at entry, in the Stanford Study of Internet Entrepreneurship. Data on the number of employees, total revenue (estimated by respondent), TechCo revenue (from administrative dataset) and income from business (estimated by respondent) were winsorized at 5% level. Firm age is calculated as the difference between the year of entry in the survey minus the reported year of the first cost incurred with the business. Respondents were excluded for not answering key questions from the survey about their business or themselves, rushing through the survey, or not being eligible to join our sample according to their responses.

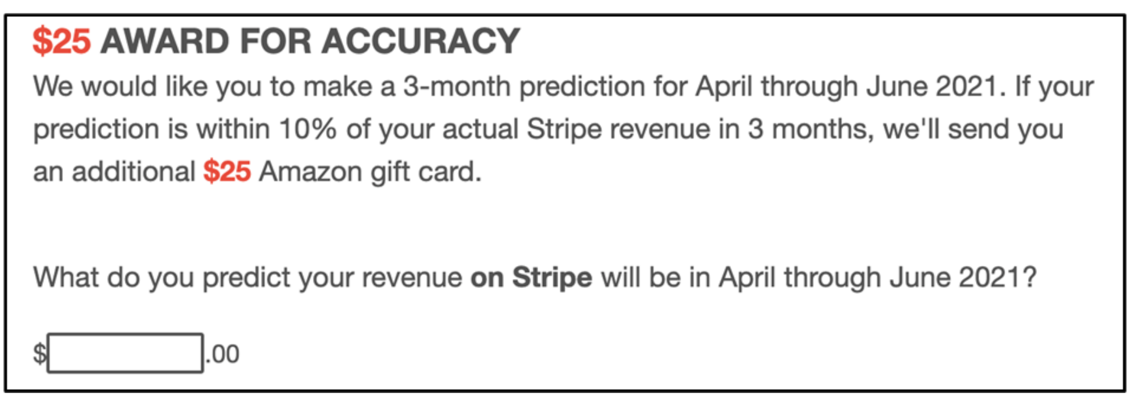
The firms span the entire United States with coverage across almost all states (as shown in Appendix Figure A.4), with most firms concentrated as expected in states such as California, Connecticut, Massachusetts, New Jersey, New York and Texas. These firms also have a broad industry mix (Appendix Figure A.1), with a skew towards industries like travel and clothing that have a higher online representation.

In terms of firm revenue characteristics, we see they are mostly small firms, with average revenue of $400,000 in the past 12 months, although there is high heterogeneity in our sample and 5% of firms have revenue in excess of $3 million at entry. The average TechCo revenue in our sample is around $150,000 in the past 12 months but the average reported percentage of revenue obtained on TechCo is just in excess of 50%, reflecting that smaller firms have higher percentage of their revenue obtained on the platform.

2.3 Prediction Competition and Reward Randomization

In each round of the survey, respondents were asked to predict their revenue on TechCo over the next three months and the next twelve months. We restricted the predictions to revenue on TechCo so that we could check their predictions directly rather than rely on reported revenue from the survey. Discussions with managers suggests platform of revenue matters due to generally big differences in fees, payout schedules, customer service, and other characteristics between payment processing competitors. For each survey, respondents were promised an Amazon gift card if their 3-month prediction was within 10% of their actual revenue. Prior to the fourth round of the survey, this was a set $25 value to ensure that respondents had a financial incentive to give credible and considered responses. The baseline question, as shown to respondents can be found in Figure 1.

Figure 1: Baseline Prediction Question


$25 AWARD FOR ACCURACY

We would like you to make a 3-month prediction for April through June 2021. If your prediction is within 10% of your actual Stripe revenue in 3 months, we'll send you an additional **$25** Amazon gift card.

What do you predict your revenue **on Stripe** will be in April through June 2021?

$ .00

In the fifth and the seventh round of the survey we randomized the reward amount that was offered to respondents. Among the respondents, 25% were randomly chosen to have no reward at all, 25% stayed with the base amount of $25, and the remaining 50% were randomly placed in eight equal sized group of 6.25% across the values of $50, $100, $150, $200, $250, $300, $350, and $400. The reward amounts were cross-randomized in each period, and were cross-randomized relative to all other treatments we applied in the surveys. The question for these surveys can be found in Appendix Figure B.1.

2.4 Payment Data

Through TechCo data we are able to track the aggregate revenue, total transactions and average transaction value for each firm directly. This is valuable in allowing us to assess survey data against businesses actual revenue data – comparing founder expectations against actuals, as well as identifying and cleaning up any major outliers. While the latter helps to reduce survey measurement error, the data do have certain limitations. Most notably, we are only able to observe the revenue that occurs on TechCo, which represents 52% of our sample's business revenue on average, according to the survey data (as can be seen in Table 2). This also means that we cannot observe revenue before the business joins TechCo, or easily distinguish between a business leaving TechCo and a business closing.[14]

On the other hand, compared to administrative data we can observe businesses before they formalize, so are able to capture information on very early-stage entrepreneurship. The TechCo transactions data is also direct revenue data, rather than data reported to tax, accounting, or statistical authorities, so is less susceptible to measurement error or misreporting.

Second, TechCo accounts are not always uniquely matched to businesses. For instance, 40% of businesses have multiple accounts. Some of these accounts may be used for testing purposes while the businesses join TechCo, while others may correspond to individual establishments owned by the business. Ideally, we could aggregate accounts, however it is not always clear which accounts belong to the same business. In some cases, founders may have multiple accounts with the same email that actually correspond to different businesses which adds additional complication. As a result, we ask for the business account TechCo code of the firm before each response in our survey.[15]

14. We address this with survey data as best we can in the results section, however, there are concerns about what fraction of businesses ultimately chose to reply to our survey after they have closed and report their closure to us.

15. We will address these issues further in subsequent work in the robustness section.

2.5 TechCo Dashboard Usage

As part of using TechCo to process their payments, businesses are given access to a financial dashboard. This dashboard, shown in Appendix Figure 4, appears immediately when users open their account. The dashboard allows users to easily see various financial metrics, including past revenue which they may reference while taking the survey and in particular while reporting past sales and predicting future sales.

We are able to observe dashboard usage passively, as each time a user accesses the dashboard it is recorded. This allows us to observe how often users access the dashboard in general, as well as see if they accessed the dashboard at the time of the survey. Despite the potential usefulness of the dashboard, there is large variation in the amount that businesses use it, as seen in Figure C.1A, which shows the number of days in 2019 that businesses used their dashboard. The average user accesses their account approximately once per month.

Our second intervention was to randomly induce respondents to look at this dashboard immediately prior to being asked for their predictions. This was done three times, in rounds five, six, and seven, and it was assigned independently of the other rounds each time. The treatment was light-touch. Respondents were simply asked to log into their TechCo accounts, confirm the account for us, and then copy their revenue over the last three months into the survey. The control group was instead asked to report their revenue from the last three months with no indication on how they should come up with this number, and they were asked to confirm the account for us after they made the prediction.

2.6 Forecasting Sales for Hypothetical Businesses

In addition to forecasts on their own business, we also asked a subset of firms to make forecasts for a series of ten hypothetical businesses in the seventh round of the survey. Respondents were first presented with a graph showing the historical revenue for a hypothetical firm and were asked to predict the revenue for the firm in the next quarter (e.g., Figure D.1). They were then shown a suggested forecast we created from an autoregressive model (e.g., Figure D.2). [16] This process was repeated ten times for the ten hypothetical firms. The or-

16. These suggested forecasts were created using an autoregressive model of next quarter's revenue on the previous four quarters (i.e, column four of table D.1). This model essentially predicts that next periods revenue will be the prior periods revenue, with only a slight weighting for the inclusion of the previous quarters.

der of the presentation of each hypothetical scenario was randomized (i.e. the first question was a random selection from the 10 possible scenarios and so on).

The inclusion of this forecasting training module was also randomized: a third of entrepreneurs were given the training module prior to being asked for their forecasts, a third of entrepreneurs were shown the module after giving their forecasts, and a third were not shown the module at all. This allows us to have a treatment group, a control group for immediate effects on predictions, and a true control group for analyzing long-term effects that show up in future rounds.

For the two-thirds of entrepreneurs who filled out the module, the module has a secondary benefit of providing a measure of forecasting ability separate from the forecasts of their own firms' sales. This allows the firms to compete on an even-playing field, whereas when forecasting their own sales many face more difficult forecasting problems due to issues such as higher volatility, less sales history to rely on, or emotional tie considerations. These issues often correlate with can correlate with performance metrics such as size and growth, so having a separate (own firm characteristics independent) measure can be useful when analyzing the connection between forecasting ability and performance.

2.7 Management

In the second round of the survey (Summer 2019), firms were asked to respond to a module on management practices. The questions mimicked with minor adjustments those from the Management and Organizational Practices Survey (MOPS) and the Annual Survey of Entrepreneurs (ASE). In total, we used 7 questions, listed in Appendix E along with their scoring. The questions cover personnel practices, the use key performance indicators, the use of targets, and the handling of issues that arise for the business.

Each question was scored on a scale of 0 to 1. The response which is associated with the most structured management practice is normalized to one, and the one associated with the least structured practices is normalized to zero. The composite management score used throughout this paper is then the simple average of those scores. Unless otherwise noted, the results presented in this paper exclude the use of the two questions on personnel practices as the majority of firms did not have enough personnel to warrant a discussion of their personnel practices. [17]

17. We later reran the management module on firms in the sixth round of the survey (Spring 2021) to

3 The State of Forecasting

We begin the analysis by analyzing the accuracy of entrepreneur forecasting. We find that even for short-run forecasting, entrepreneurs make both highly inaccurate guesses and are on average over-optimistic. We find that sporadic sales tracking on the platforms provided by TechCo (including the TechCo Dashboard) is very strongly negatively correlated with forecast accuracy.

3.1 Entrepreneurs Forecast Poorly

Despite the presumably significant importance of forecasting for businesses, we find that entrepreneurs perform poorly when asked to make forecasts. To evaluate forecast accuracy graphically and in our models, we calculate error as follows[18]:

$$ForecastError_{it} = log(Predict_{it}) - log(Actual_{it}) \tag{1}$$

Figure 2 plots a histogram of forecast errors of the 3-months exercise for the periods prior to the introduction of the interventions.[19] Entrepreneurs in these pre-periods were all offered \$25 rewards for forecasts that were within 10% of the realized value. This 10% window is overlayed on top of the histogram for reference. Even with the reward, we find that only 13% of forecasts are within the window.

Graphically, it is apparent that the errors have a wide variance with little asymmetry. Using the following means squared error (MSE) decomposition:

$$MSE = Bias^2 + Variance \tag{2}$$

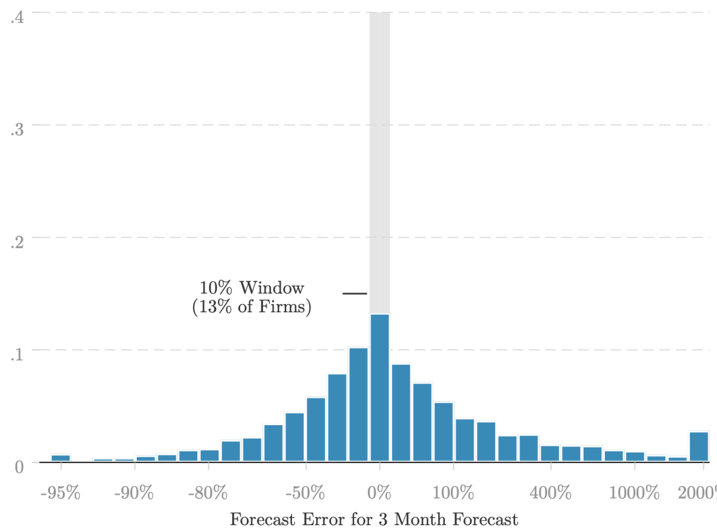
In Table 3 we show that 92.7% of the aggregate mean squared error is noise (variance), and only 7.3% is bias. While the entrepreneurs are moderately overconfident, the errors they make are dominated by non-systematic noise. To benchmark this estimate, we consider a simple random walk model that inputs the sales from the previous three months as the prediction. This model would achieve a 2.05 pp. higher win rate and a massive reduction in Means Squared Errors, of just above a quarter.

gather panel management data on the firms who remained in the survey and gather new management data on the firms who had entered through refresh samples or had separately missed the second round.

18. Of course, win rates are calculated based on variables that are not log-transformed, as announced to survey participants.

19. We decided to drop round 4, as predictions were made for the start of 2020 which was heavily impacted by COVID-19. More information about this round can be found in Bloom, Fletcher, and Yeh 2021

Figure 2: Forecast Accuracy in 3-months prediction exercise



Note: Forecasting error is calculated as log(forecast next quarter sales) – log(realization of next quarter sales). Results for rounds 1-3, 5300 firms. All firms were paid $25 for quarterly sales forecasts within 10% of their actual numbers.

Table 3: Baseline win rates, mean squared error decomposition, random walk benchmark

	Baseline	Random Walk	Difference	% Difference
Win rate	12.98%	15.02%	+2.04pp.	-
MSE	0.892	0.662	-0.231	100.0%
Bias^2	0.065	0.001	-0.064	27.7%
Noise	0.827	0.661	-0.167	72.3%

Notes: Data comes from 5,300 firms and 10,017 survey responses, in the Stanford Study of Internet Entrepreneurship, rounds 1-3. Win rates calculated using 10% from actual sales definition. MSE decomposition calculated using Equation 2. Random walk model inputs the sales from the previous three months as the prediction.

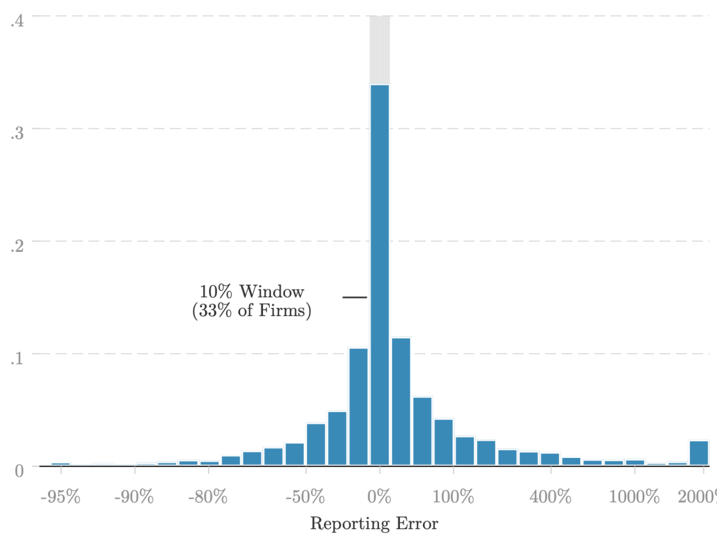
3.2 Entrepreneurs Report Revenue Poorly

A major factor underlying the poor performance of the entrepreneurs' forecasts is the underlying inaccuracy of previous revenue reporting. To analyze reporting accuracy, we calculate error as follows:

$$ReportError_{i,t} = log(Reported\_Prev_{i,t}) - log(Actual\_Prev_{i,t})$$

Figure 3 plots a histogram of reporting errors for the previous 3 months, periods prior to the introduction of the interventions. For direct comparison with the forecast errors in Figure 2, we again overlay a 10% window on top of the histogram. Despite the fact that reporting past revenue is a significantly easier endeavor, we still find that that only around 33% of revenue reports are within the window.

Figure 3: Historical Sales Reporting Accuracy for previous 3-months



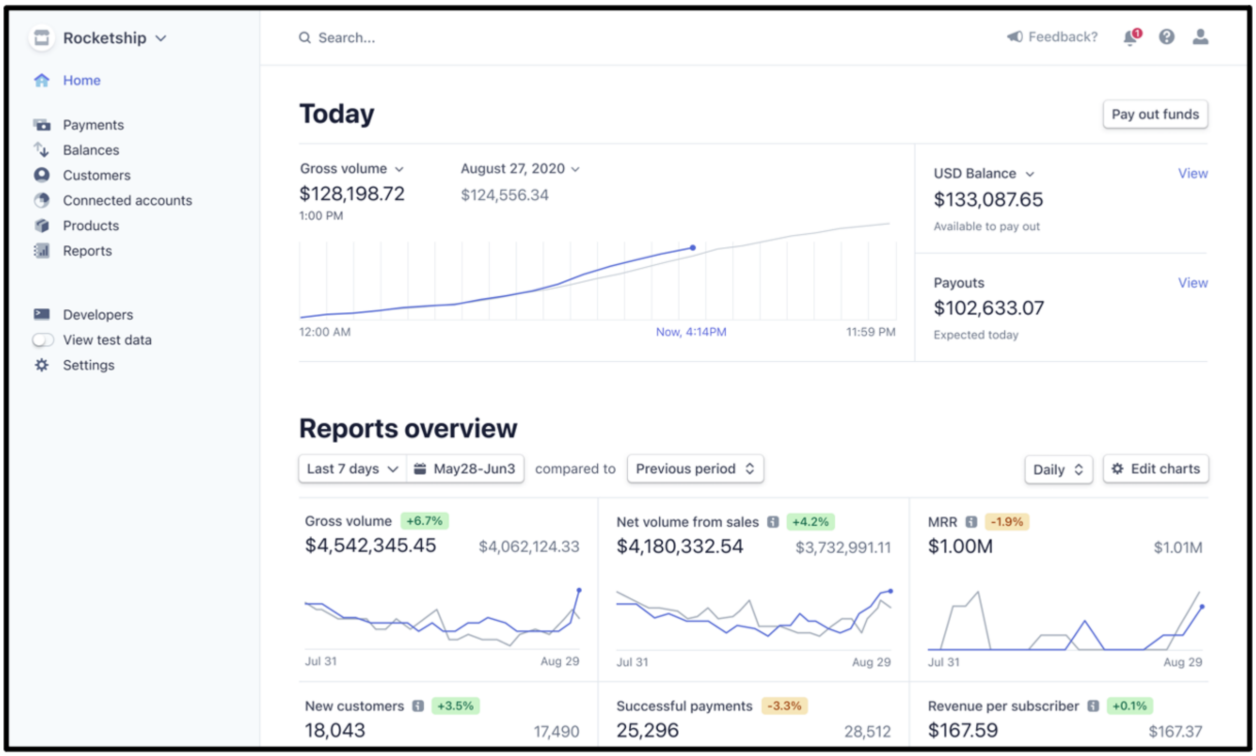
Note: Reporting error is calculated as log(reported last quarter sales) – log(last quarter sales). Results for rounds 1-3, 5300 firms.

3.3 Dashboard Use and Accuracy

The inaccuracy of (the non-incentivised) past revenue reporting for almost 70% of responses is especially troubling considering the ease with which this information is available. As mentioned in the first section, using the dashboard is an easy way for businesses to consult their financial data for the purposes of reporting revenue and forecasting future sales. Given that users are able to easily consult their past revenue, it begs the question why respondents are performing so poorly in reporting their past revenue and forecasting their future revenue. The Dashboard is presented in Figure 4.

Certainly, our results appear to show that dashboard can be very useful. Figure C.1A shows a binned scatter plot of forecast errors versus the number of times that firm used the dashboard in 2019. We indeed see that daily users of the dashboard perform significantly better.

Figure 4: TechCo Dashboard for a generic firm



Beyond general use of the dashboard and general awareness of financial status, we can also disaggregate the results by use of the dashboard during the survey. Figures C.1B and C.1C show the results for forecasting and reporting error respectively by whether or not individuals viewed their dashboard the day that they completed the survey. In Figure C.1B we see that forecast errors are much smaller when respondents used the dashboard. The difference is even more striking when it comes to reporting past revenue accurately (Figure C.1C). Firms that use the dashboard are significantly more likely to accurately report their revenue.

4 Can Forecasts be Improved?

We now turn to our second central research question: what types of treatments best suit the types of errors entrepreneurs make? We analyze this question using three classifications of possible intervention: increasing attention, improving the quality of data and information used as a basis for the forecast, and improving the forecaster's ability to better leverage information.

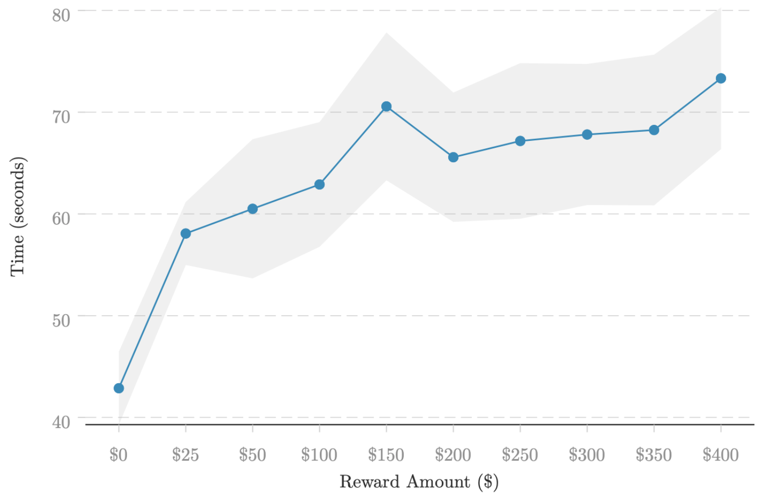
4.1 Effects of Higher Reward Payments

Our first intervention was to the randomization of reward amounts for correct forecasts. We first document that the payments were successful in inducing entrepreneurs to spend more time answering the prediction exercise. We then show that while this extra time and effort did result in a decreased level of average over-optimism but it did not significantly reduce the larger issue of noise.

Figure 5 plots non-parametrically the time spent predicting sales for the next 3-months. The relationship is clearly positive, with the amount of time increasing steadily with the reward amount offered for a correct forecast. The shape is approximately concave, with the biggest effects occurring with the first \$25 as entrepreneurs jump from an average of 42 seconds to nearly 60 seconds. After this initial jump, the amount begins to peter out, with an additional \$375 only adding approximately 15 seconds, which is just about 25% of the time spent at \$25 reward levels. We interpret this as suggesting the while money is effective in getting individuals to spend more time on the survey, respondents quickly run out of ways to spend that additional time and do not know how to move much beyond the basic forecast they provide with minimal reward. Table 4 shows the regression equivalent of Figure 5. In column 1, we see that each additional \$100 that we added to the reward led on average to an additional 5.8 seconds of time spent. Columns 2 and 3 show a robustness check that dashboard treatment does not have a significant effect and the estimate in column 1 is stable.

In addition to the increased time spent forecasting, we also see that respondents are less over-optimistic when they are paid greater rewards for accuracy. Figure B.2 shows that bias decreases with the amount paid, and respondents who are paid offered the maximum reward are almost unbiased. However, the treatment does not meaningfully affect overall

Figure 5: Time Spent Forecasting Vs. Reward Amount ($)



Notes: Time took to answer the 3-months forecasting exercise. Times are winsorized at 180 seconds, and trimmed to drop respondents who went through the survey in times too short to have read and comprehended the questions. Sample of 3,177 firms from round 5 and 7.

Table 4: Prediction Timing by Treatment

	Time (s) Prediction	Time (s) Prediction	Time (s) Prediction
Reward '00s	5.762***		5.763***
	(0.778)		(0.778)
Dash Treat		-0.406	-0.462
		(1.562)	(1.545)
Dep. Mean	56.369	56.369	56.369
Observations	3177	3177	3177

Notes: Time took to answer the 3-months forecasting exercise. Times are winsorized at 180 seconds, and trimmed to drop respondents who went through the survey in times too short to have read and comprehended the questions. Sample of 3,177 firms from round 5 and 7.

mean squared error. Figure B.3 graphically highlights that there is no clear relationship between reward amount and the mean squared forecast error. It is possible for bias to fall while mean squared error is not significantly effected given that noise is the overwhelmingly larger driver of mean squared error.

Table 5 shows the regressions matching Figures B.2 and B.3. In column 1 and 2, we see the regressions of report error and report error squared on the reward amount. This acts as a placebo test since past revenue is reported before the reward amount is revealed. Encouragingly, there is no significant effect. Column 3 shows the effect of the reward on forecast error, which is significantly negative 3.4 percentage point reduction in bias. Column four looks at the squared forecast error, and has a negative point estimate but is not significant as already foreshadowed.

Table 5: Reward Treatment Effects

	Report Err.	(Report Err.)^2	Forecast Err.	(Forecast Err.)^2
Reward '00s	-0.008	0.002	-0.034**	-0.032
	(0.010)	(0.022)	(0.014)	(0.026)
Time FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Dep. Mean	0.030	0.414	0.127	0.768
Observations	6659	6659	6659	6659

Notes: Regression of log of (forecast next quarter sales) – (realization of next quarter sales) on the reward payment for forecasts within 10% of actual. Data from rounds 1 to 7, with standard errors clustered at the firm level.

All together, this suggests that rewards can induce entrepreneurs to spend longer time and approach their forecasts more realistically. This has the effect of lowering over-optimism. The intervention does not give them better information and does not teach them to better use they information they do have, and so does not lead to lower total mean squared error. While entrepreneurs can be induced to think more realistically, the results suggest that entrepreneurs are not aware of how to improve their forecasts very much and quickly reach their best forecast.

4.2 Effects of Dashboard Usage

Turning to the dashboard treatment, we test whether better information, and in particular historical sales data, can significantly reduce forecasting error. To test this, we look for effects of showing the dashboard to entrepreneurs first on reporting accuracy for historical sales before turning to forecast accuracy. We find that both reporting and forecast accuracy are significantly improved by the intervention.

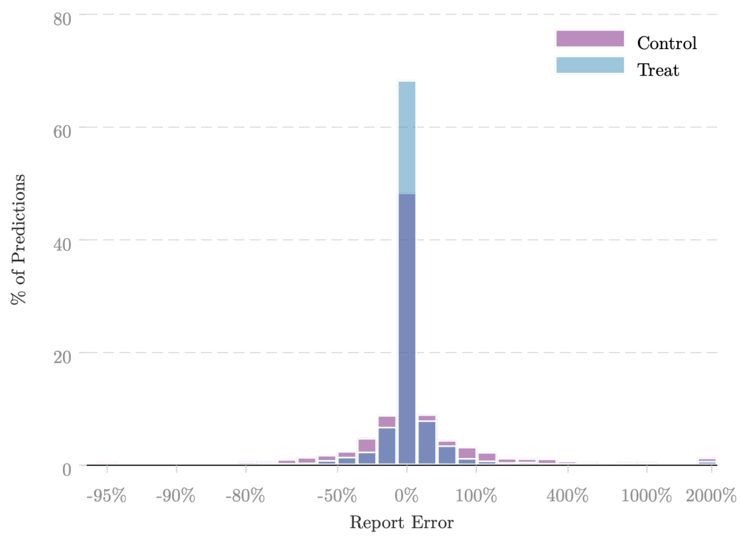
We begin by showing the significant effect of the dashboard treatment on both the reported error and the absolute reporting error. Figure 6 shows the reduction is quite large. This effect is not surprising in the least, given that respondents are in instructed to report directly from their dashboard in the treatment group. It does, however, confirm that individuals who are misreporting their revenue can, and by and large have access to this information.

Table 6 performs the same regressions as Table 5 with the Dashboard treatment instead. The first two columns of Table 6 shows the result of the dashboard treatment on the entrepreneurs' reporting errors. We find that there is a 2.2 percentage points reduction in reporting error, and a 21.7 percentage points reduction in squared reporting error. Columns 4 show that this leads to a corresponding reduction in squared forecast error of 11.4 percentage points. Column 3 shows that the bias on forecasts is not statistically significantly reduced, although the point estimate is negative (at face value, it also shows a reduction of around 10% compared to the control group). We interpret this as evidence that the reduction in bias of reporting error likely does lead to a reduction in the bias in forecast error, but that a larger driver of bias in forecast error may be other factors such as realism as discussed in the context of the reward treatment. In the Appendix, Tables C.1-C.3 show the effects of the dashboard intervention by the share of revenue on TechCo, total revenue and dashboard usage. The effects are generally higher for firms with lower TechCo usage (below median) and those with lower revenue (below median).

4.3 Effects of Training

Last of all we examine the training treatment. To test whether the training had an effect, we look for learning withing the training itself, before turning to forecast accuracy. While we find significant learning within the training module itself, the accrued knowledge does not translate to changes in forecasting accuracy. We find that both forecasting errors and

Figure 6: Reporting Error (past 3-months) by Dashboard Treatment



Notes: Reporting error for the last quarter is calculated as log(reported last quarter sales) – log(last quarter sales). Sample of 3,975 firms from round 5 to 7.

Table 6: Dashboard Treatment Effect

	Report Err.	(Report Err.)^2	Forecast Err.	(Forecast Err.)^2
Dashboard	-0.022	-0.217***	-0.012	-0.114**
	(0.022)	(0.042)	(0.029)	(0.054)
Time FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Dep. Mean	0.107	0.695	0.099	1.034
Observations	6659	6659	6659	6659

Notes: Regression of log of (forecast next quarter sales) – log (realization of next quarter sales) on the dashboard treatment for forecasts within 10% of actual. Data from rounds 1 through 7, with standard errors clustered at the firm level.

squared forecast errors are not significantly affected.

4.3.1 Learning

The 10 questions in the training module were randomly ordered for each respondent, so we can evaluate learning throughout the training by evaluating the average response accuracy on the first question seen vs. the second question seen vs. the third question seen and so on. If no learning occured, then there will be no improvement in the average response accuracy over time. The difficulty, however, is that for these training forecasts were made on hypothetical data, so there is no correct amount that is realized over time and there is not a precise way to measure accuracy. Instead, we rely on three alternative measures of performance.

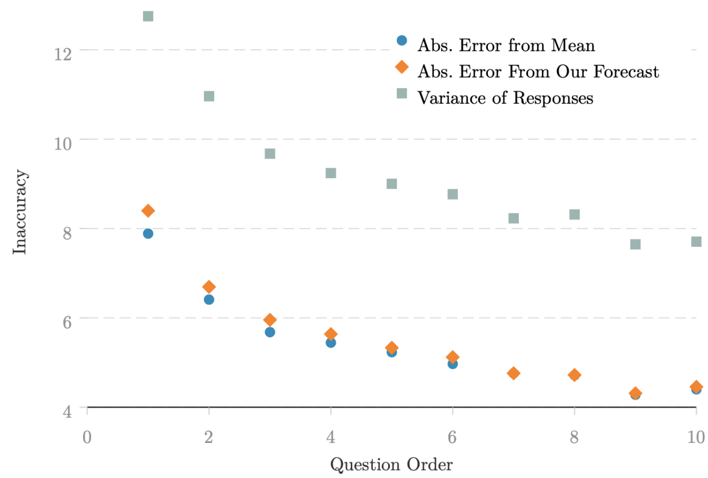
First, we can compare the forecasts entrepreneurs make to the suggested forecasts we show them after they make each forecast and before they move to the next. These forecasts are, again, based on a simple autoregressive model of next quarter's revenue on the previous four quarters (Table D.1). While this metric relies in part on the idea that our forecasts are reasonable, they are being encouraged to match their forecasts to ours and the module is teaching them how to make forecasts similar to these. So as long as better forecasters start nearer on average to our forecasts or converge quicker, the measure is meaningful.

Second, we can compare forecasts to the mean forecast for each question. This has the advantage of not needing a potentially arbitrary choice of forecasting method to compare them to and is generally valid so long as there is some 'wisdom of the crowd'.

Last of all, we can evaluate the overall variance of forecasts. As can been seen in the bias variance decomposition, mean squared error increases directly with variance (i.e., there is more noise and so forecasts are worse). This metric of performance does not rely on having a "correct" realization that forecasts are aiming for as it does not concern itself with the bias component of the error.

Figure 7 shows how individuals performed on each question using all three measures. For each metric, there is a monotonic improvement in performance with question they see that levels off by the 10th question at the end of the training module. This demonstrates clear learning within the the module.

Figure 7: Response Accuracy by Question Order in the Training



Note: Absolute error from the mean is calculated as the absolute difference between the forecast and the average forecast. Variance of responses is the variance of all responses for a given question number. Survey participants answered all 10 questions in a random order. Data from 2,953 firms in Round 7.

4.3.2 Effects of Training on Actual Forecasts

Unlike the previous two interventions, we find no statistically significant at conventional levels effect at all from the training on either forecast error or square forecast error.

The first two columns of Table 7 shows the result of the training treatment on the entrepreneurs' forecast errors (which can be considered a placebo, just like in the reward treatment). We find that there is a no significant change in reporting error or in squared reporting error. Columns 3 and 4 show the effect on forecasting error and squared forecasting error, which is again insignificant in both cases but negative in sign. Figures D.3 and D.4 graphically demonstrate the same null effect for the forecast error and forecast error squared, respectively.

These null effects are indication that either training is a generally ineffective way to increase forecast accuracy or our particular training was ineffective. While our training was light-touch in nature, we do believe the results are reflective of training overall. First, there is strong evidence of learning within the training module itself and our forecasting methods significantly overperform the predictions made by the respondents. Second, forecasting sales

Table 7: Training Treatment Effect

	Report Err.	(Report Err.)^2	Forecast Err.	(Forecast Err.)^2
Training	-0.012	0.053	-0.029	-0.020
	(0.041)	(0.081)	(0.055)	(0.097)
Time FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Dep. Mean	0.096	0.592	0.090	0.946
Observations	6659	6659	6659	6659

Regression of log of (forecast next quarter sales) – log (realization of next quarter sales) on the dashboard treatment for forecasts within 10% of actual. Data from rounds 1 through 7, with standard errors clustered at the firm level.

for small businesses is difficult as the individual circumstances (that the entrepreneurs have more information on) tend to be very heterogeneous across specific businesses and industries. Therefore, gains in accuracy beyond using historical data to establish base rates can be difficult to achieve.

4.4 Comparison of Effects

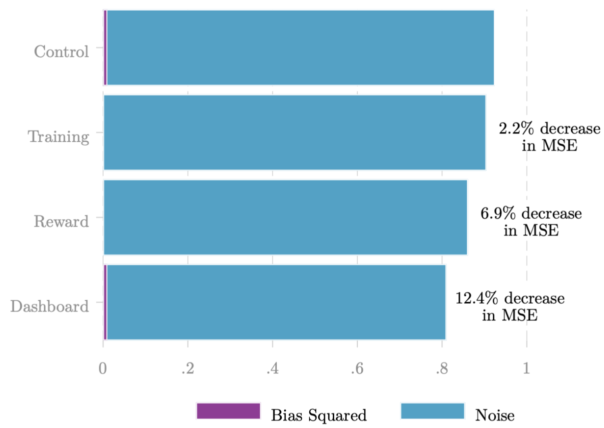
Figure 9 aggregates all of the previous results to cleanly show the magnitudes and natures of each of the three treatments. In each of the treatment arms, noise is the overwhelming source of errors, and a treatment that reduces noise is virtually guaranteed to be the most effective.[20]

Starting with the training treatment, we found that it had only a small 2.2% effect on mean squared error. This is because it had little effect on both bias or noise reduction[21]. The reward treatment has slightly a slightly larger 6.9% impact of as it does have some bias reduction, even if the noise reduction is minimal. In plain English, the reward treatment

20. Note the difference in the control group in terms of MSE noise/bias decomposition and the previous baseline figures. This is caused by the specifics of the waves 5-7, and possibly the effect of the Covid-19 shock, reflecting over-optimism as a smaller share of MSE.

21. Noise reduction is calculated using the decomposition of mean squared error into bias squared and noise, where the $Bias^2$ reduction is calculated for a respondent with average bias and an average reduction in bias due to treatment.

Figure 8: Comparison of treatments



Note: Data is from rounds 5 through 7, with 3,975 unique firms. Reward effects are calculated for the average payment value in our experiment of $200.

improves accuracy by making managers more realistic, but the improvement is still small because it does nothing to improve the overall variability of their forecasts.

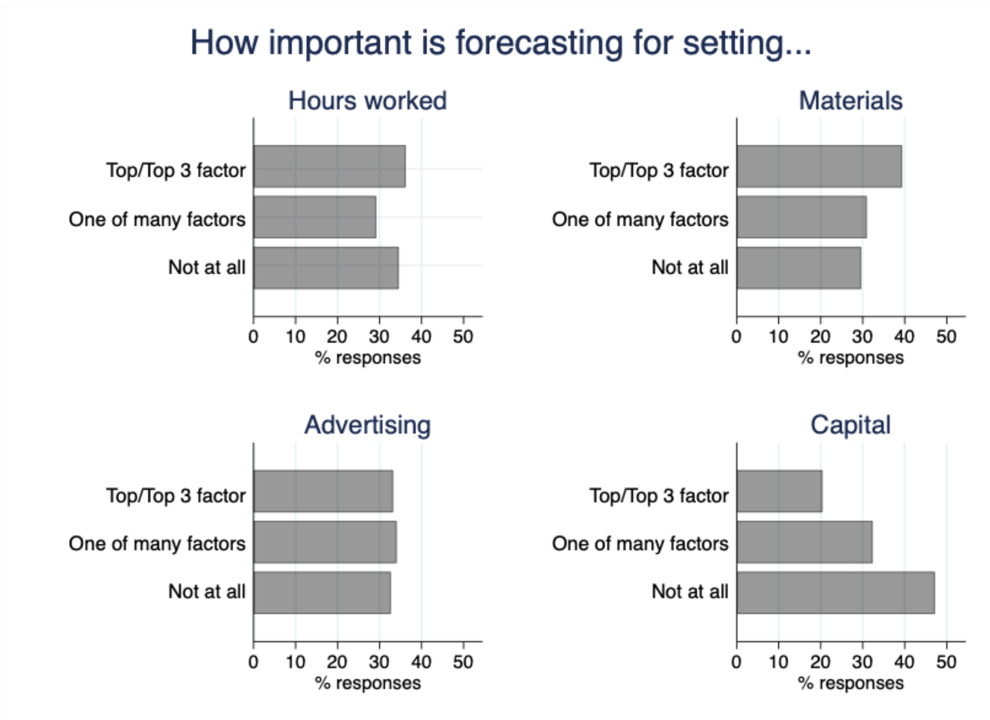
The most effective treatment comes from the dashboard intervention. While there was little effect on bias, the intervention was an extremely effective way of decreasing noise in managerial forecasts. Given the importance of noise as a driver of error, it was able to reduce MSE by 12.4%, which is almost twice the point estimate effect of the average reward amount.

5 The Importance of Forecasting

Given our results that entrepreneurs perform forecasts poorly despite the availability of sales data that can significantly improve their accuracy, a natural concern is that forecasting does not matter for entrepreneurs. While we do not believe this to be the case, we do want to address it directly as it would directly undermine the purpose of the study.

First, we directly ask entrepreneurs how important sales forecasting (such as the 3-months and 12-months exercises we have performed) are for setting various inputs in their production processes. Forecasting sales seems to be most important for setting materials expenditures and hours worked, with around 40% of the sample naming it as a top/top 3 factor. Forecasting sales does not seem to be a very important input for setting capital expenditures, suggesting that capital may be set as a result of a longer-run/forward looking analysis, or more financial constraints may affect its level. Nonetheless, around 90% of sample report that sales forecasts are a top/top 3 factor for deciding at least one of hours worked, number of employees, material, capital or advertising expenditures. In Appendix Figure A.11, we offer more quantitative information about the self-reported changes in inputs between worst and best sales scenarios.

Figure 9: Self-reported importance of forecasting for setting various inputs



Note: Data is from round 8, for the first 1,640 participating firms.

Second, as already discussed above, we offer quite substantial rewards for accurate predictions. These rewards had no effect on forecast accuracy despite the fact that the entrepreneurs could easily have improved their accuracy with minimal effort by accessing the dashboard. The fact that they did not do so is evidence that even when the incentives are clear, they aren't aware of the benefits of data-driven forecasting and so, at least in the short-term, failure to adopt these methods is not driven by a lack of value but by a lack of knowledge.

Third, we have substantial evidence linking forecasting performance with firm performance. This evidence we present here extends to both forecasting for their own firm and forecasting for hypothetical firms that abstract from the intricacies of their own situation.

5.1 Forecasting and Firm Performance

We find that forecasting accuracy is significantly associated with business performance on all of main metrics of business performance: revenue, growth, survival, and a composite management score.

Figure 10A shows that firms with higher accuracy are bigger, as measured by quarterly sales from the quarter immediately prior to the forecast period. Figure 10B shows that firms with higher accuracy grow faster, where growth is measured as the percentage change in revenue in the period following the forecast. Figure 10C shows that firms with higher accuracy are more likely to survive (in the TechCo dataset). As discussed in the first section, survival is hard to measure because we cannot perfectly distinguish between firms who close and firms who choose to no longer use TechCo.[22] As a proxy of survival, we use whether or not a firm has had sales in the last 6 months but this finding is robust to adjusting the period considered.[23] Last of all, Figure 10D shows that forecast accuracy is positively associated with our composite management score described in section 2. This is not surprising as the management score reflects the use of KPIs, recording of data, and general business sophistication, all of which can relate directly to forecast performance.

5.2 Forecasting Training Module and Firm Performance

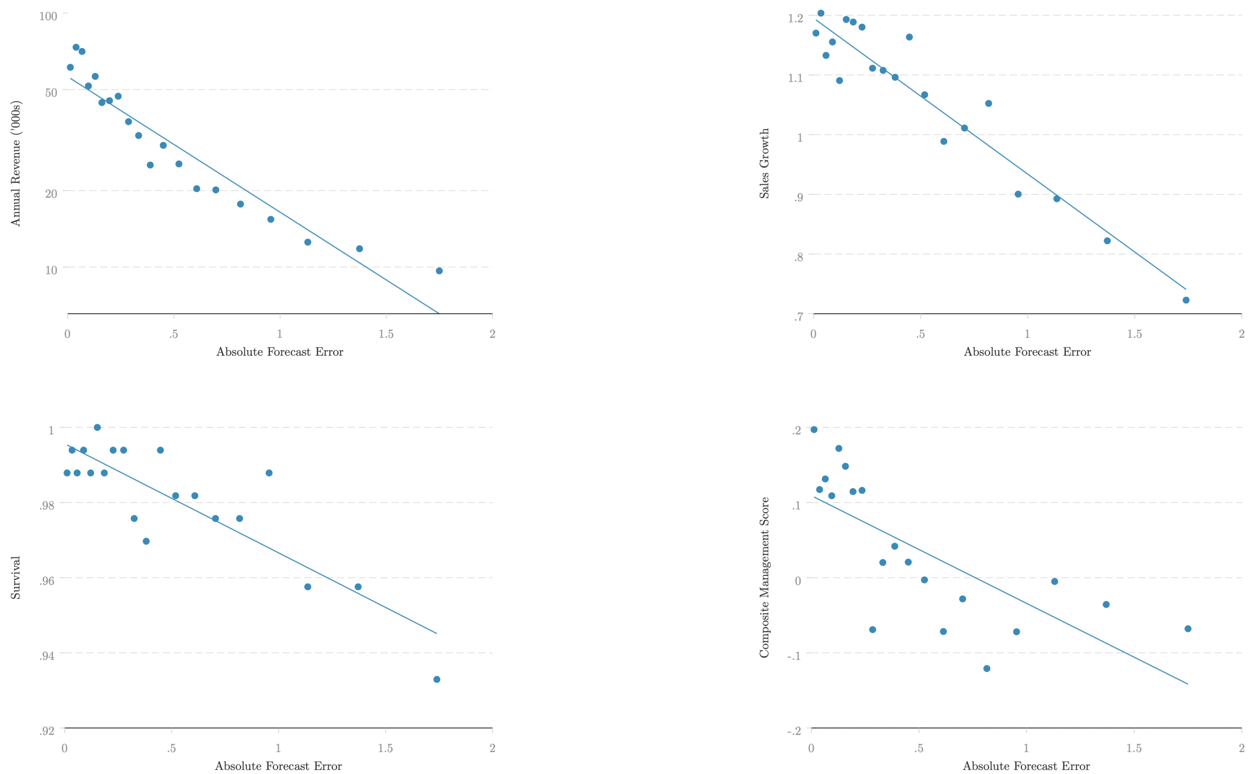
In addition to forecasts on their own business, we can also use performance on the training module as a measure of forecasting skill. The performance on the training module has the advantage that is consistent across all firms and not affected to the underlying sales process each firm faces. As evidence that it is indeed picking of forecasting ability, Figure 11 shows that there is a strong positive correlation between forecasting errors on the training module and the actual forecasting errors for the firms.

We can now compare performance on the forecasting module with firm outcomes just

22. We ask firms to report whether or not they have closed in the survey, but there is likely much larger attrition from the survey for closed firms which would bias survival estimates downwards.

23. Some firms also go through long periods of time greater than 6 months with no sales before they reemerge with sales, so this does falsely estimate some firms as closed when they are still in business.

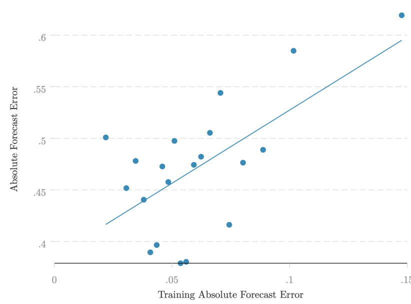
Figure 10: Performance metrics and forecasting accuracy



Note: Annual registered TechCo firm revenue, quarterly growth calculated using TechCo data, semi-annual survival rates in TechCo data and Composite Management Score (calculated as mentioned in the main text), plotted against annual absolute forecast error in 3-months prediction exercise. Historical data from rounds 1 through 7, for 6,659 participating firms.

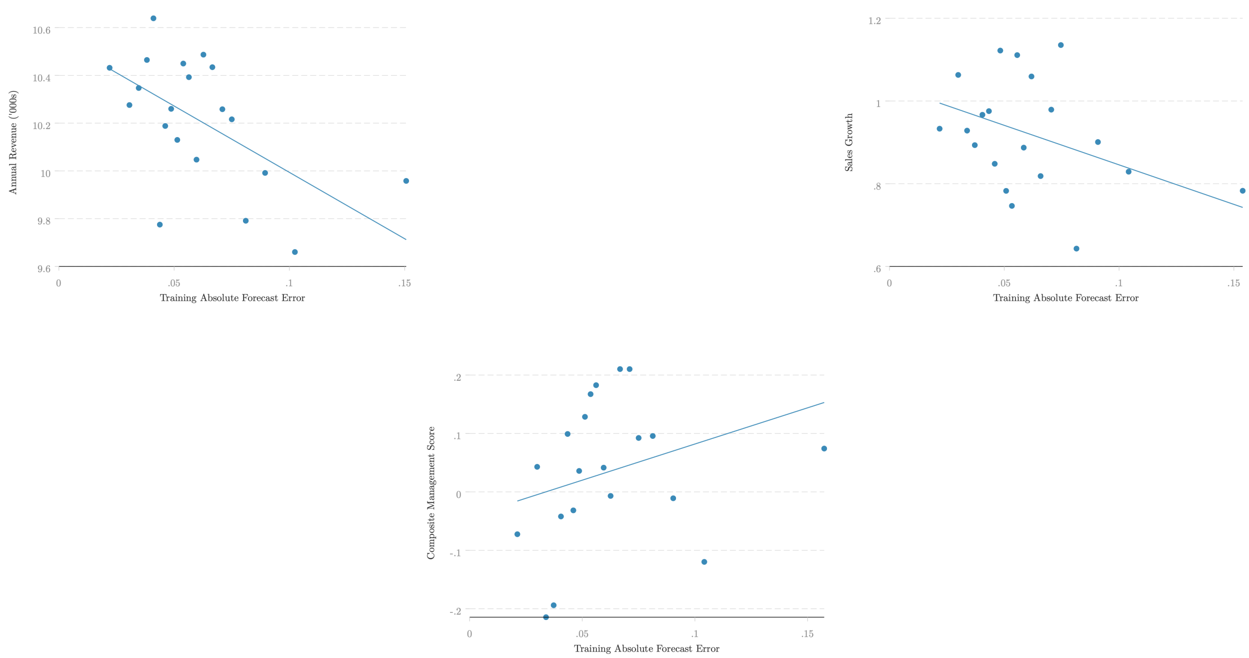
as before. Notably, we find the same strong relationships between forecasting and firm performance. Figure 12A shows that firms who perform better in the training are larger. They are also better managed and have higher growth rates (Figure 12B, Figure 12C).

Figure 11: Forecasting Accuracy Vs. Training Forecasting Accuracy



Note: Forecast error are historical data from rounds 1 through 7. Training forecast errors are from the training module in round 7.

Figure 12: Performance vs. Training Forecasting Accuracy



Note: Annual registered TechCo firm revenue, quarterly growth calculated using TechCo data, semi-annual survival rates in TechCo data and Composite Management Score (calculated as mentioned in the main text). All are historical data from rounds 1 through 7, for 6,659 participating firms. Training forecast errors are from the training module in round 7.

6 Why Don't Forecasts Improve?

Given that entrepreneurs perform poorly when forecasting, that forecasts are monetary incentivized (and they seem to be correlated with various firm performance metrics so they "matter"), and that they can easily improve their forecasts using readily available data, the obvious question is why they do not do so. We believe that the answer lies with general overconfidence and an inability to assess their own relative forecasting abilities. These mistakes can cause them to undervalue the benefits of adopting a data-driven forecasting.

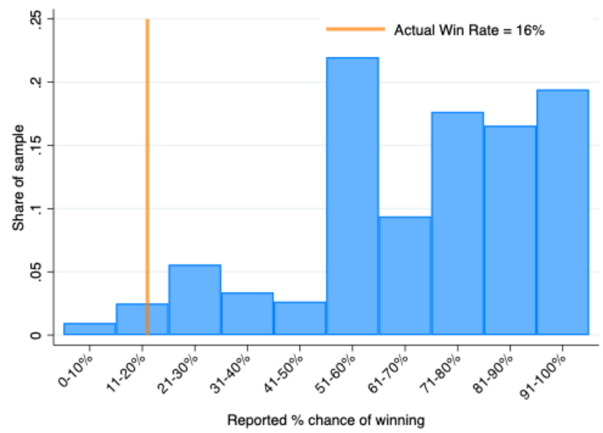
Nowhere is the overconfidence clearer than in Figure 13. In the most recent completed round of the survey, immediately after making their predictions, respondents were asked what probability they had of winning. Despite an average win rate of 19.3%, the vast majority report winning odds of 50% or greater.

Overconfidence is also evident from the ranges that respondents report when asked for their best case and worst case forecast for sales. The realized sales value is outside of the range of worst to best cases a striking 38.8% of the time. Respondents are therefore significantly overconfident in their forecasts.

On top of the general overconfidence we observe, respondents also have difficulty in reporting their own relative forecasting ability. Before respondents completed the forecast training module, which again provides a standard measure of forecasting ability, we asked respondents to report their self-assessed forecasting ability on a 5-point likert scale going from far below average to far above average. Assigning each category a respective value from 1 to 5, with 1 being far below average and 5 being far above average, we can compare reported ability with actual ability. Figure 14 shows a binscatter of this self-reported forecasting ability against average absolute error on the training questions. There is a strong relationship, with those who had larger errors reporting themselves as the best.

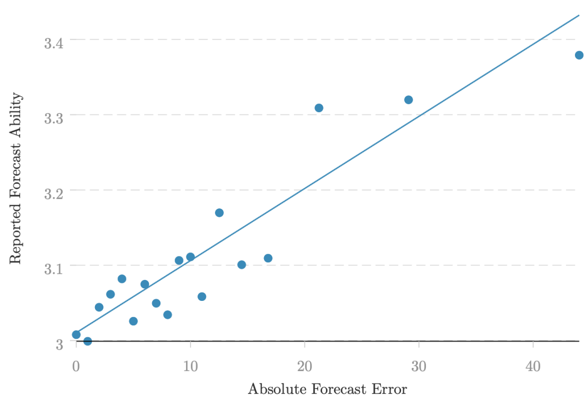
Taken together, these results demonstrate that individuals are not accurately gauging the quality of their forecasting and so likely are not properly valuing the value of additional tools to help themselves forecast.

Figure 13: Reported Odds of Winning Forecast Contest (%)



Note: Reported probabilities of winning were collected in Round 8. The average win rate was calculated for the first 986 firms in Round 8.

Figure 14: Reported Forecasting Ability Vs. Measured Performance



Note: Self-reported abilities were collected in round 7 using a 5-point likert scale with far below average as 1 and far above average as 5. The absolute error is calculated using the absolute difference between their response and the suggested response in the training module, from 2,953 firms in Round 7.

7 Conclusions

We have gathered data on manager expectations for future sales using an ongoing survey of online firms in partnership with TechCo, an online payment processing company. In each round of the survey, managers are asked to predict future sales, which we can compare directly with their actual account data. At baseline, only 13% of firms can forecast their sales within 10% of the realized value, with 5% of the error attributable to bias and the remaining 95% attributable to noise.

We investigate three possible types of interventions to help improve forecast accuracy: increasing attention, improving the quality of data and information used as a basis for the forecast, and improving the forecaster's ability to better leverage information. For each of these types, we design and experimentally evaluate a specific intervention.

Our first intervention increases attention by rewarding entrepreneurs up to $400 for forecasts within 10% of the realized value. Our second intervention improves the quality of data used in forecasts by instructing respondents to review their historical sales data prior to making their forecast. Our results from these two interventions suggest that the reward payment significantly reduces bias but has no effect on noise. The historical sales data intervention has no effect on bias but significantly reduces noise. Since bias is only a minor part of overall forecasting errors, we find that the reward payments have negligible effects on mean squared error, while the historical data intervention reduces mean squared error by 14%. These results suggest that while paying firms results in more realistic forecasts, firms benefit more through the use of data-driven forecasting and decision making.

As mentioned above, there is a lingering question of why firms do not make greater use of their readily available dashboards and other sources of historical sales performance for their forecasts given the benefits. We provide evidence that it is not because forecasting accuracy isn't valuable to them, but instead because they are inaccurate in their beliefs about their own skill and likely because they undervalue the data available on their dashboard. In terms of policy implications, this suggests that the most effective way of influencing firm expectations (for monetary and fiscal policymakers) may be increasing access to past performance data and encouraging the use of historical data-driven forecasting.

References

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2022. "Management and Shocks to Worker Productivity." *Journal of Political Economy* 130, no. 1 (January): 1–47. https://doi.org/10.1086/717046. https://doi.org/10.1086/717046.

Alekseev, Georgij, Safaa Amer, Manasa Gopal, Theresa Kuchler, JW Schneider, Johannes Stroebel, and Nils Wernerfelt. 2020. *The Effects of COVID-19 on U.S. Small Businesses: Evidence from Owners, Managers, and Employees.* Technical report. September. https://doi.org/10.3386/w27833. https://doi.org/10.3386/w27833.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Brent H. Meyer, and Nicholas Parker. 2019. *Surveying Business Uncertainty.* Working Paper 25956. National Bureau of Economic Research, June. Accessed October 24, 2021. https://doi.org/10.3386/w25956. https://www.nber.org/papers/w25956.

Anderson, Stephen J., and David McKenzie. 2022. "Improving Business Practices and the Boundary of the Entrepreneur: A Randomized Experiment Comparing Training, Consulting, Insourcing, and Outsourcing." *Journal of Political Economy* 130, no. 1 (January): 157–209. https://doi.org/10.1086/717044. https://doi.org/10.1086/717044.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation." Publisher: The University of Chicago Press, *Journal of Political Economy* 122, no. 5 (October): 1013–1063. ISSN: 0022-3808, accessed October 24, 2021. https://doi.org/10.1086/677072. https://www.journals.uchicago.edu/doi/abs/10.1086/677072.

Bachmann, Rüdiger, Kai Carstensen, Stefan Lautenbacher, and Martin Schneider. 2020. "Uncertainty and Change: Survey Evidence of Firms' Subjective Beliefs" [in en], 86.

Bachmann, Rüdiger, and Steffen Elstner. 2015. "Firm optimism and pessimism." *European Economic Review* 79 (October): 297–325. https://doi.org/10.1016/j.euroecorev.2015.07.017. https://doi.org/10.1016/j.euroecorev.2015.07.017.

Barrero, Jose Maria. 2021. "The micro and macro of managerial beliefs" [in en]. *Journal of Financial Economics* (June). ISSN: 0304-405X, accessed October 24, 2021. https://doi.org/10.1016/j.jfineco.2021.06.007. https://www.sciencedirect.com/science/article/pii/S0304405X21002737.

Bartik, Alexander W., Marianne Bertrand, Zoe Cullen, Edward L. Glaeser, Michael Luca, and Christopher Stanton. 2020. "The impact of COVID-19 on small business outcomes and expectations." *Proceedings of the National Academy of Sciences* 117, no. 30 (July): 17656–17666. https://doi.org/10.1073/pnas.2006991117. https://doi.org/10.1073/pnas.2006991117.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. 2013. "Managerial Miscalibration*" [in en]. Publisher: Oxford Academic, *The Quarterly Journal of Economics* 128, no. 4 (November): 1547–1584. ISSN: 0033-5533, accessed May 13, 2020. https://doi.org/10.1093/qje/qjt023. https://academic-oup-com.stanford.idm.oclc.org/qje/article/128/4/1547/1850166.

Bertrand, M., and A. Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics* 118, no. 4 (November): 1169–1208. https://doi.org/10.1162/003355303322552775. https://doi.org/10.1162/003355303322552775.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What Drives Differences in Management Practices?" *American Economic Review* 109, no. 5 (May): 1648–1683. https://doi.org/10.1257/aer.20170491. https://doi.org/10.1257/aer.20170491.

Bloom, Nicholas, Robert Fletcher, and Ethan Yeh. 2021. *The Impact of COVID-19 on US Firms.* Technical report. January. https://doi.org/10.3386/w28314. https://doi.org/10.3386/w28314.

Bruhin, Adrian, Luiés Santos-Pinto, and David Staubli. 2018. "How do beliefs about skill affect risky decisions?" *Journal of Economic Behavior & Organization* 150 (June): 350–371. https://doi.org/10.1016/j.jebo.2018.01.016. https://doi.org/10.1016/j.jebo.2018.01.016.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2018. "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico." *Journal of Political Economy* 126, no. 2 (April): 635–687. https://doi.org/10.1086/696154. https://doi.org/10.1086/696154.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about Politics" [in English]. Publisher: Now Publishers, Inc. *Quarterly Journal of Political Science* 10, no. 4 (December): 519–578. ISSN: 1554-0626, 1554-0634, accessed December 20, 2021. https://doi.org/10.1561/100.00014074. https://www.nowpublishers.com/article/Details/QJPS-14074.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" [In en]. *Journal of Political Economy* 120, no. 1 (February): 116–159. ISSN: 0022-3808, 1537-534X, accessed December 21, 2021. https://doi.org/10.1086/665662. https://www.journals.uchicago.edu/doi/10.1086/665662.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2020. "Inflation Expectations and Firm Decisions: New Causal Evidence*." *The Quarterly Journal of Economics* 135, no. 1 (February): 165–219. ISSN: 0033-5533, accessed December 21, 2021. https://doi.org/10.1093/qje/qjz029. https://doi.org/10.1093/qje/qjz029.

David, Joel M., and Venky Venkateswaran. 2019. "The Sources of Capital Misallocation" [in en]. *American Economic Review* 109, no. 7 (July): 2531–2567. ISSN: 0002-8282, accessed December 20, 2021. https://doi.org/10.1257/aer.20180336. http://www.aeaweb.org/articles?id=10.1257/aer.20180336.

Galasso, Alberto, and Timothy S. Simcoe. 2011. "CEO Overconfidence and Innovation." *Management Science* 57, no. 8 (August): 1469–1484. https://doi.org/10.1287/mnsc.1110.1374. https://doi.org/10.1287/mnsc.1110.1374.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer. 2016. "Expectations and Investment." Publisher: The University of Chicago Press, *NBER Macroeconomics Annual* 30 (January): 379–431. ISSN: 0889-3365, accessed December 20, 2021. https://doi.org/10.1086/685965. http://www.journals.uchicago.edu/doi/full/10.1086/685965.

Giorcelli, Michela. 2019. "The Long-Term Effects of Management and Technology Transfers." *American Economic Review* 109, no. 1 (January): 121–152. https://doi.org/10.1257/aer.20170619. https://doi.org/10.1257/aer.20170619.

Goel, Anand M., and Anjan V. Thakor. 2008. "Overconfidence, CEO Selection, and Corporate Governance." *The Journal of Finance* 63, no. 6 (November): 2737–2784. https://doi.org/10.1111/j.1540-6261.2008.01412.x. https://doi.org/10.1111/j.1540-6261.2008.01412.x.

Gosnell, Greer K., John A. List, and Robert D. Metcalfe. 2020. "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains." *Journal of Political Economy* 128, no. 4 (April): 1195–1233. https://doi.org/10.1086/705375. https://doi.org/10.1086/705375.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica* 50, no. 1 (January): 213. https://doi.org/10.2307/1912538. https://doi.org/10.2307/1912538.

Hebert, Camille. 2021. *Learning from Errors in Entrepreneurship* [in en]. SSRN Scholarly Paper ID 3947926. Rochester, NY: Social Science Research Network, November. Accessed December 21, 2021. https://doi.org/10.2139/ssrn.3947926. https://papers.ssrn.com/abstract=3947926.

Hirshleifer, David, Angie Low, and Siew Hong Teoh. 2012. "Are Overconfident CEOs Better Innovators?" *The Journal of Finance* 67, no. 4 (July): 1457–1498. https://doi.org/10.1111/j.1540-6261.2012.01753.x. https://doi.org/10.1111/j.1540-6261.2012.01753.x.

Hoffman, Mitchell, and Steven Tadelis. 2021. "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis." *Journal of Political Economy* 129, no. 1 (January): 243–285. https://doi.org/10.1086/711409. https://doi.org/10.1086/711409.

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. 2021. *Noise: a flaw in human judgment.* Little, Brown.

Larkin, Ian, and Stephen Leider. 2012. "Incentive Schemes, Sorting, and Behavioral Biases of Employees: Experimental Evidence." *American Economic Journal: Microeconomics* 4, no. 2 (May): 184–214. https://doi.org/10.1257/mic.4.2.184. https://doi.org/10.1257/mic.4.2.184.

Ma, Yueran, David Sraer, and David Thesmar. 2020. "Do Managerial Forecasting Biases Matter?" [In en], 39.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*." *The Quarterly Journal of Economics* 126, no. 1 (February): 373–416. ISSN: 0033-5533, accessed December 21, 2021. https://doi.org/10.1093/qje/qjq004. https://doi.org/10.1093/qje/qjq004.

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *The Journal of Finance* 60, no. 6 (November): 2661–2700. https://doi.org/10.1111/j.1540-6261.2005.00813.x. https://doi.org/10.1111/j.1540-6261.2005.00813.x.

———. 2008. "Who makes acquisitions? CEO overconfidence and the market's reaction." *Journal of Financial Economics* 89, no. 1 (July): 20–43. https://doi.org/10.1016/j.jfineco.2007.07.002. https://doi.org/10.1016/j.jfineco.2007.07.002.

Malmendier, Ulrike, and Geoffrey Tate. 2015. "Behavioral CEOs: The Role of Managerial Overconfidence" [in en]. *Journal of Economic Perspectives* 29, no. 4 (November): 37–60. ISSN: 0895-3309, accessed December 21, 2021. https://doi.org/10.1257/jep.29.4.37. http://www.aeaweb.org/articles?id=10.1257/jep.29.4.37.

Massenot, Baptiste, and Yuri Pettinicchi. 2018. "Can firms see into the future? Survey evidence from Germany." *Journal of Economic Behavior & Organization* 145 (January): 66–79. https://doi.org/10.1016/j.jebo.2017.10.018. https://doi.org/10.1016/j.jebo.2017.10.018.

Mellers, Barbara, Lyle Ungar, Jonathan Baron, Jaime Ramos, Burcu Gurcay, Katrina Fincher, Sydney E. Scott, Don Moore, Pavel Atanasov, Samuel A. Swift, Terry Murray, Eric Stone, and Philip E. Tetlock. 2014. "Psychological Strategies for Winning a Geopolitical Forecasting Tournament" [in en]. Publisher: SAGE Publications Inc, *Psychological Science* 25, no. 5 (May): 1106–1115. ISSN: 0956-7976, accessed October 27, 2021. https://doi.org/10.1177/0956797614524255. https://doi.org/10.1177/0956797614524255.

Moore, Don A., and Paul J. Healy. 2008. "The trouble with overconfidence." *Psychological Review* 115, no. 2 (April): 502–517. https://doi.org/10.1037/0033-295x.115.2.502. https://doi.org/10.1037/0033-295x.115.2.502.

Santos-Pinto, Luis, and Leonidas Enrique de la Rosa. 2020. "Overconfidence in Labor Markets." In *Handbook of Labor, Human Resources and Population Economics,* edited by Klaus F. Zimmermann, 1–42. Cham: Springer International Publishing. ISBN: 978-3-319-57365-6. https://doi.org/10.1007/978-3-319-57365-6_117-1. https://doi.org/10.1007/978-3-319-57365-6_117-1.

Satopää, Ville, Marat Salikhov, Philip Tetlock, and Barb Mellers. 2021. *Decomposing the Effects of Crowd-Wisdom Aggregators: The Bias-Information-Noise (BIN) Model* [in en]. SSRN Scholarly Paper ID 3781405. Rochester, NY: Social Science Research Network, February. Accessed October 27, 2021. https://doi.org/10.2139/ssrn.3781405. https://papers.ssrn.com/abstract=3781405.

Tanaka, Mari, Nicholas Bloom, Joel M. David, and Maiko Koga. 2019. "Firm performance and macro forecast accuracy" [in en]. *Journal of Monetary Economics* (March). ISSN: 0304-3932, accessed May 12, 2020. https://doi.org/10.1016/j.jmoneco.2019.02.008. http://www.sciencedirect.com/science/article/pii/S0304393219300467.

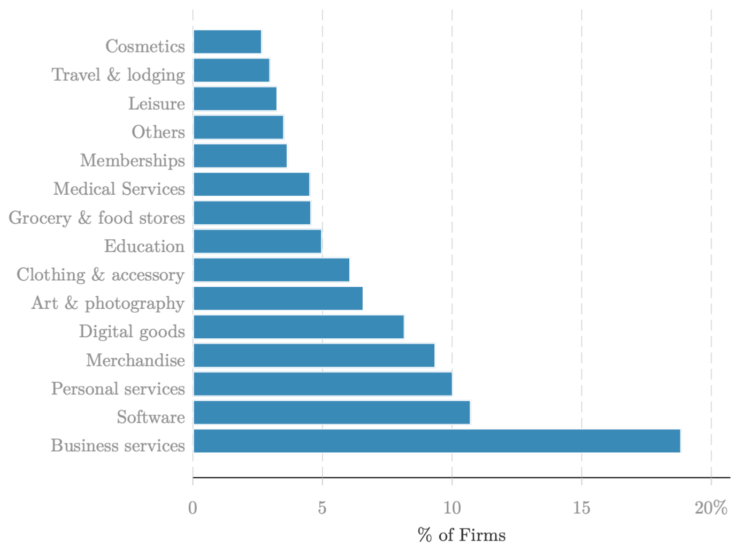
A Additional Details

Table A.1: Propensity to Respond

	Finished	Finished	Finished	Finished
Log Revenue	-0.003***	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Funded		-0.048***	-0.046***	-0.044***
		(0.006)	(0.006)	(0.006)
Industry FEs			Yes	Yes
Region FEs				Yes
F-test Industry			0.000	0.000
F-test Region				0.001
R2	0.001	0.003	0.009	0.010
Adj R2	0.000	0.003	0.008	0.008
Dep. Mean	0.227	0.227	0.227	0.227
# Obs	23069	23069	23069	23060

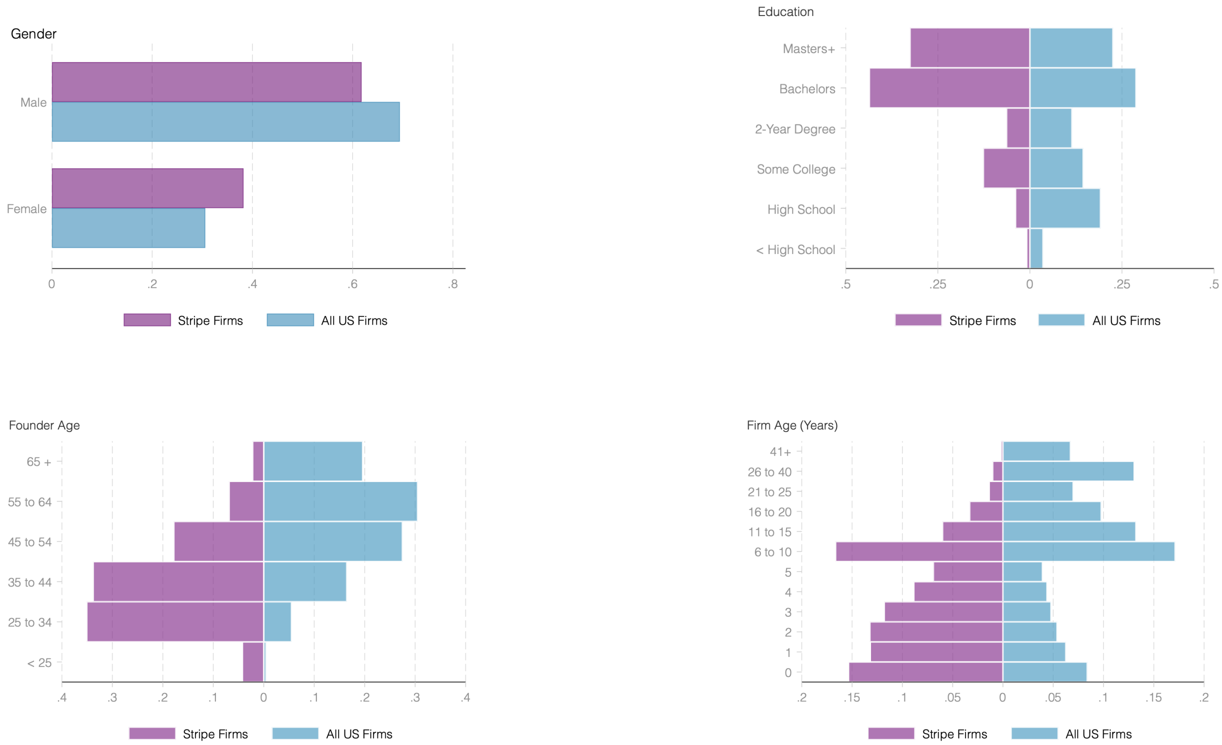
Notes: Data for firms comes from 7,463 survey respondents in the Stanford-TechCo Study of Internet Entrepreneurship. Finishing corresponds with ever completing (and sending) a survey response.

Figure A.1: Industries



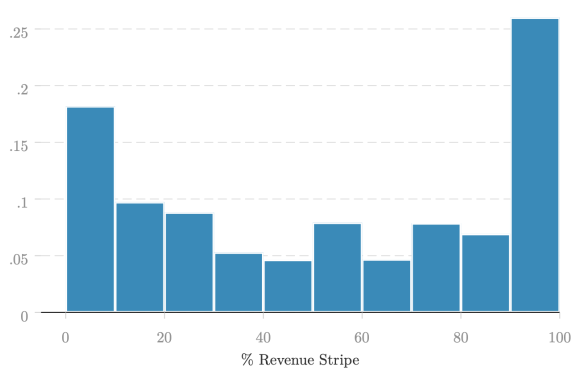
Note: Data for online firms comes from 7,463 survey responses on the Stanford-TechCo Study of Internet Entrepreneurship.

Figure A.2: Comparison of Sample Users Vs. U.S. Businesses



Data on the U.S. businesses aggregates comes from the Annual Survey of Entrepreneurs 2019 (henceforth ASE), which is a nationally representative survey of all (rather than TechCo) businesses.

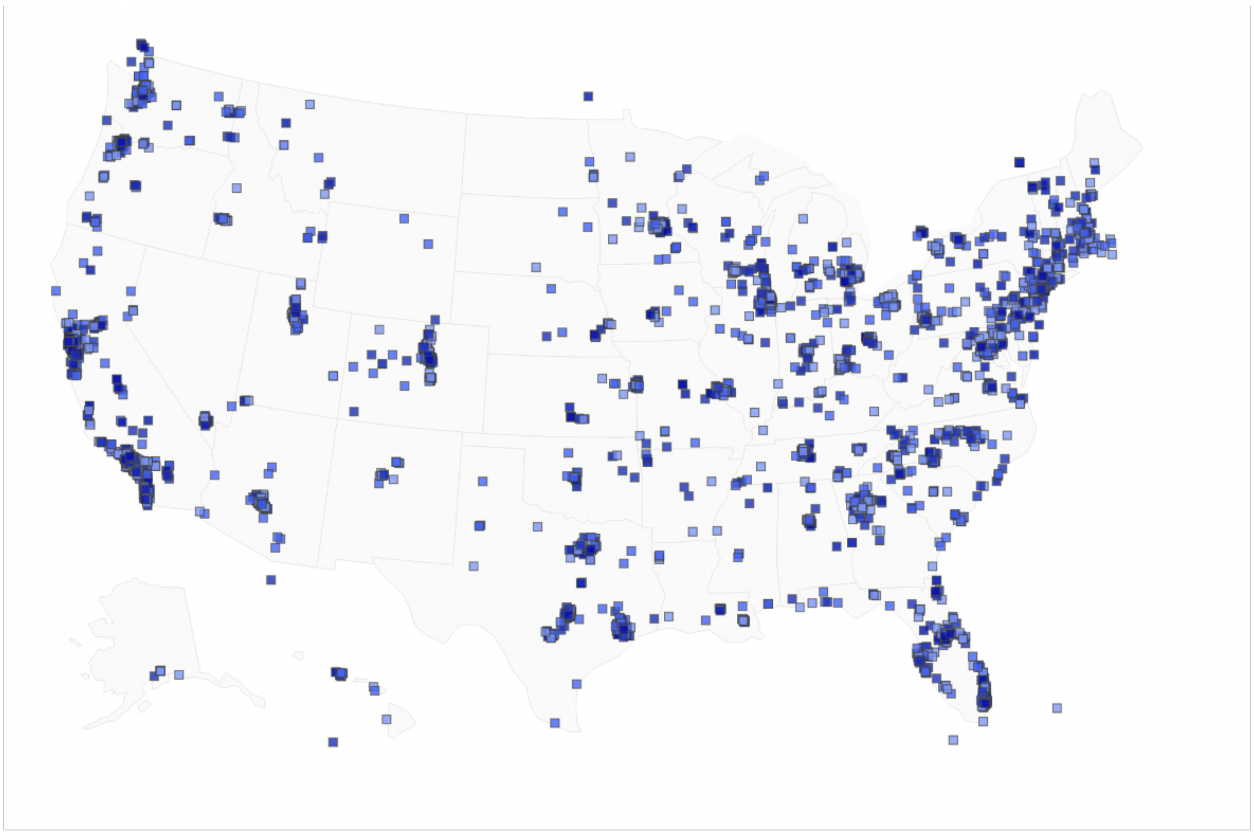
Figure A.3: Reported % of Revenue on TechCo



Note: Data for online firms comes from survey responses to the Stanford-Stripe Study of Internet Entrepreneurship.

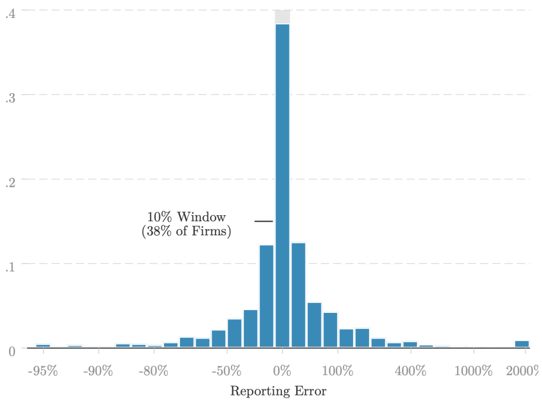
Note: Reflecting firms answering in Rounds 1-7, data is self-reported.

Figure A.4: Geography of Businesses



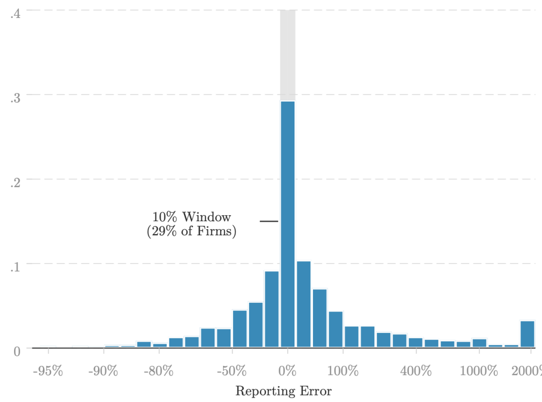
Note: Data for online firms comes from 7,463 survey responses on the Stanford-TechCo Study of Internet Entrepreneurship

Figure A.5: Reporting Accuracy by Business Type



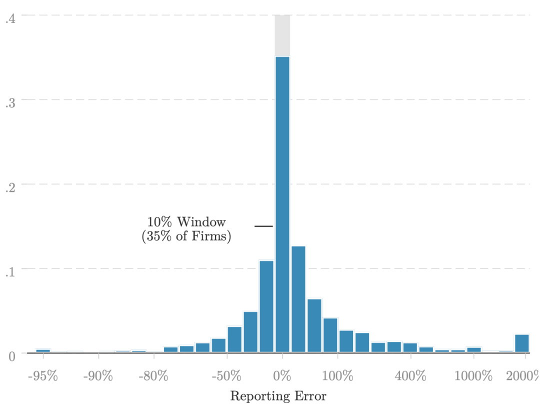
(a) Big Business

Note: Reported sales accuracy for responses in the big business strata (higher revenue than median)



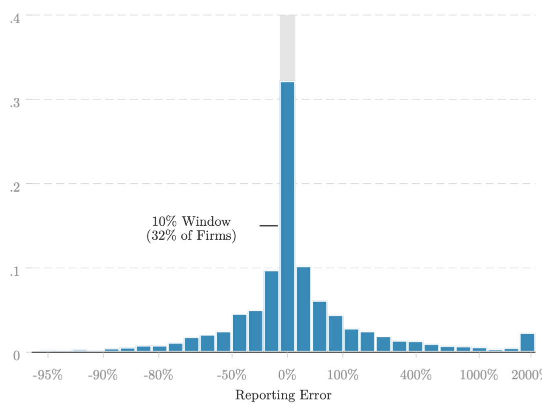
(b) Small Business

Note: Reported sales accuracy for responses in the small business strata (lower revenue than median)



(c) High TechCo User

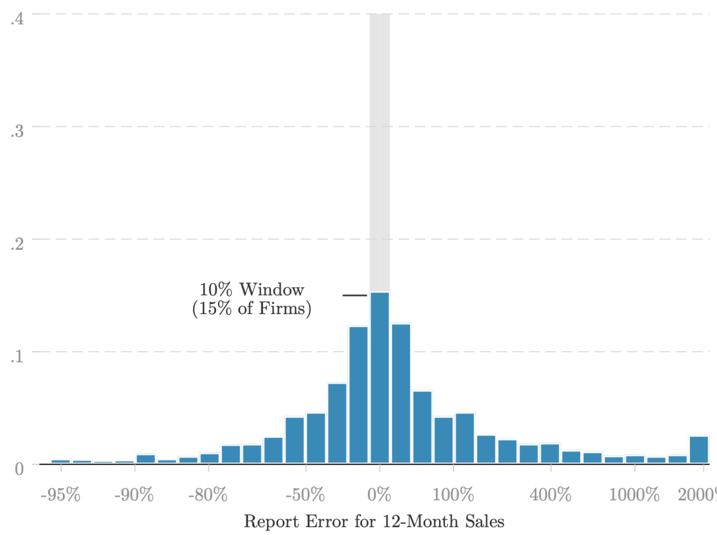
Note: Reported sales accuracy for responses with greater than 50% of their sales revenue on TechCo



(d) Low TechCo User

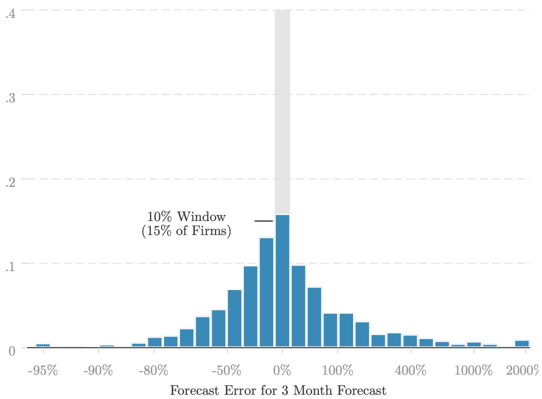
Note: Reported sales accuracy for responses with less than 50% of their sales revenue on TechCo

Figure A.6: Reporting accuracy for 12-Months, baseline



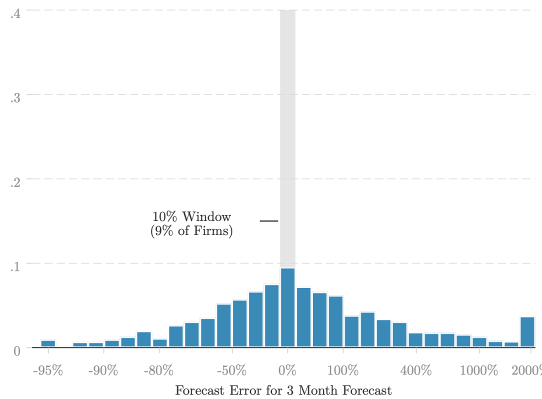
Note: Reporting error is calculated as log(reported last year sales) – log(last year sales). Results for rounds 1-3, 5300 firms.

Figure A.7: Forecasting Accuracy by Business Type



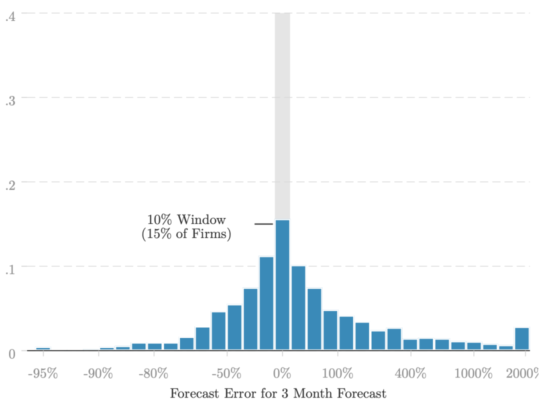
(a) Big Business

Note: Forecasted sales accuracy for responses in the big business strata (higher revenue than median)



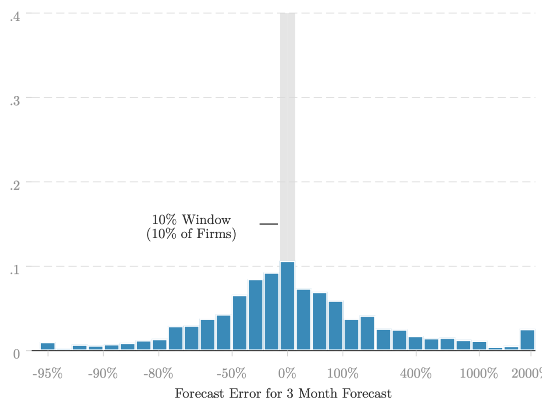
(b) Small Business

Note: Forecasted sales accuracy for responses in the small business strata (lower revenue than median)



(c) High TechCo User

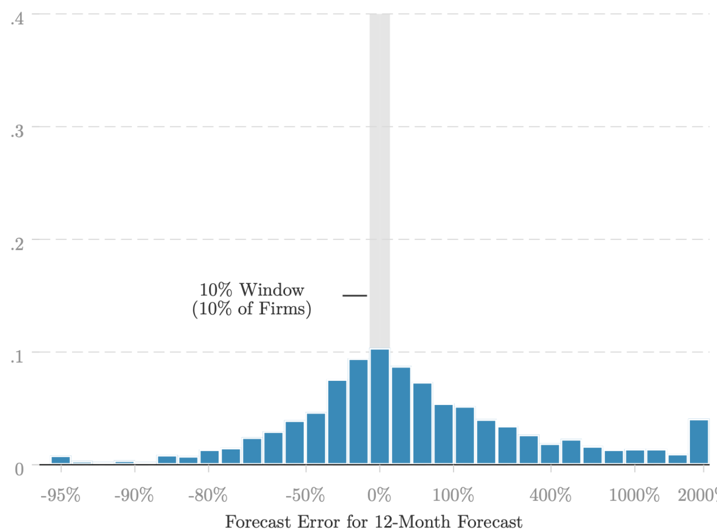
Note: Forecasted sales accuracy for responses with greater than 50% of their sales revenue on TechCo



(d) Low TechCo User

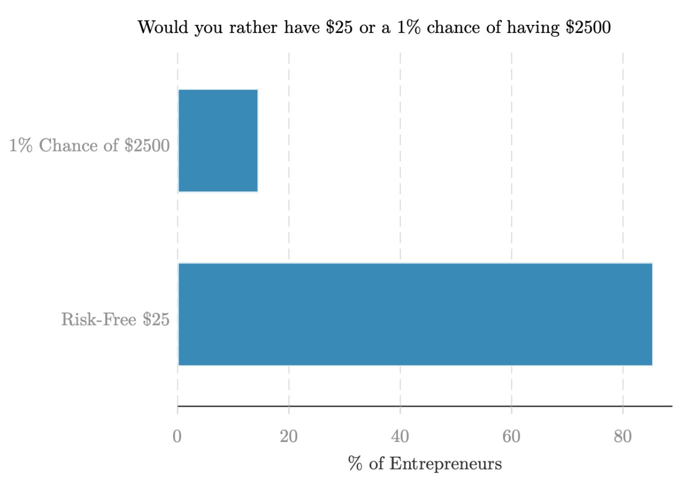
Note: Forecasted sales accuracy for responses with less than 50% of their sales revenue on TechCo

Figure A.8: Forecasting accuracy for 12-Months isn't any better



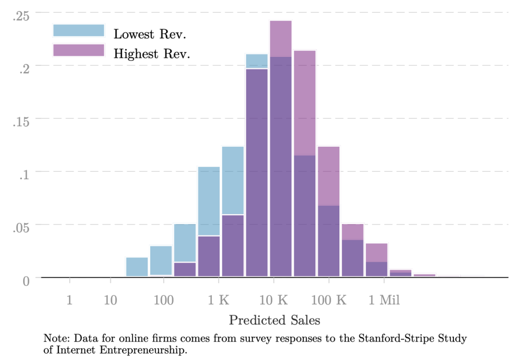
Note: Forecasting error is calculated as log(forecast next year's sales) – log(realization of next year's sales). Results for rounds 1-3, 5300 firms.

Figure A.9: Entrepreneurs are not risk loving



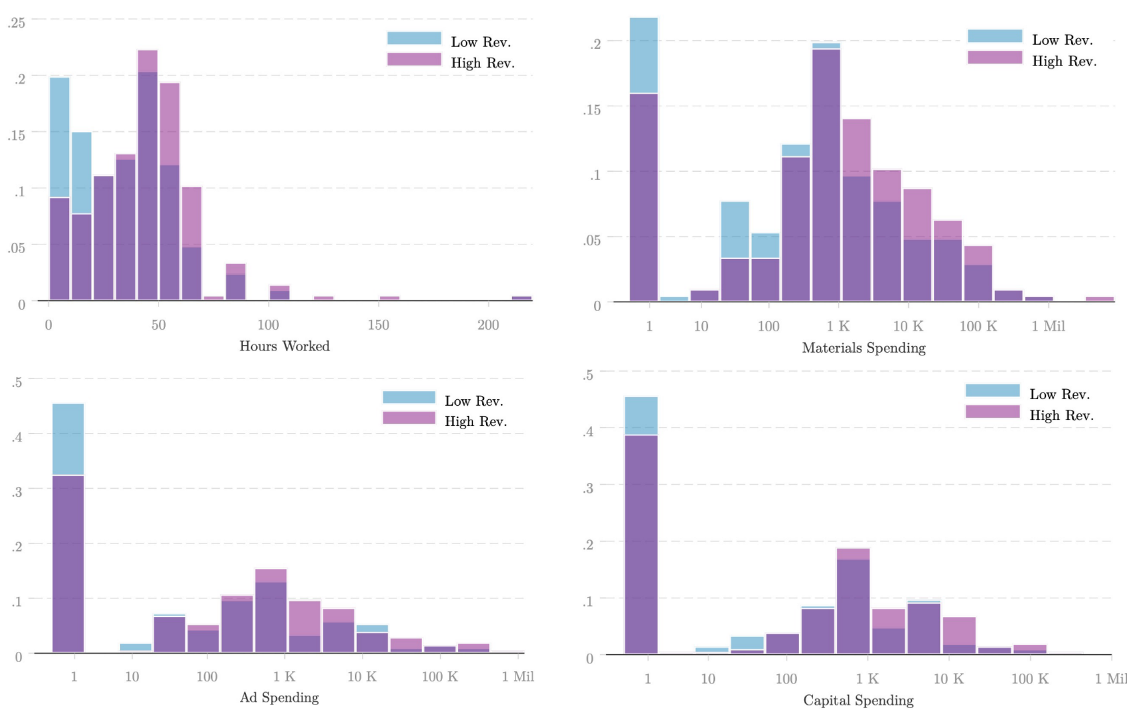
Note: Risk aversion is induced from bet choice reports in round 3 of the survey.

Figure A.10: Highest and lowest sales predictions



Notes: Changes between self-reported highest and lowest sales for the next quarter scenarios. Only first 1,640 firms participating in wave 8.

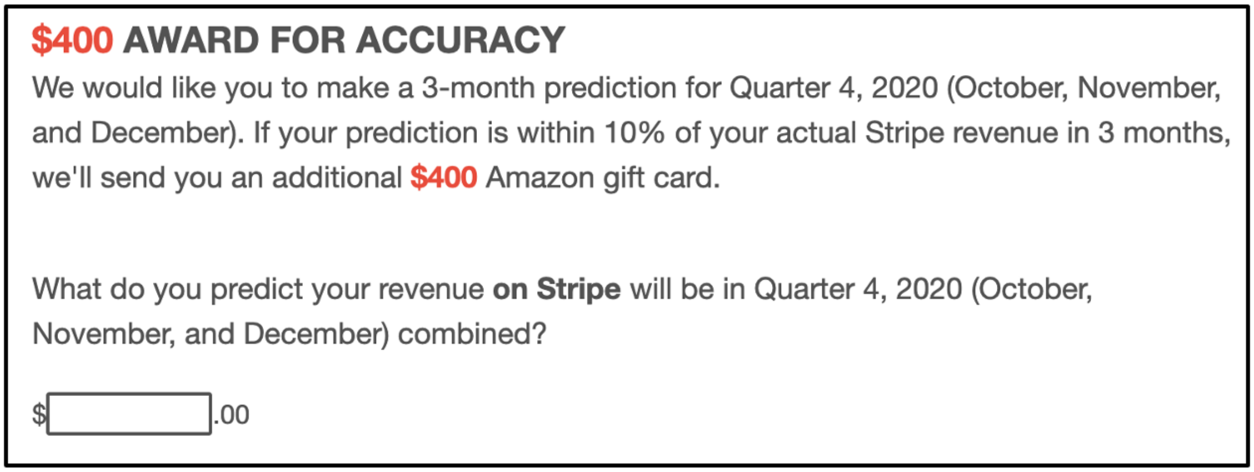
Figure A.11: Changes in inputs in different states



Notes: Changes between self-reported highest and lowest next quarter sales scenarios for hours worked (ULS) and materials (URS), advertising (DLS) and capital (DRS) expenditures. Only first 1,640 firms participating in wave 8.

B Reward

Figure B.1: Prediction Question



$400 AWARD FOR ACCURACY

We would like you to make a 3-month prediction for Quarter 4, 2020 (October, November, and December). If your prediction is within 10% of your actual Stripe revenue in 3 months, we'll send you an additional **$400** Amazon gift card.

What do you predict your revenue **on Stripe** will be in Quarter 4, 2020 (October, November, and December) combined?

$ .00

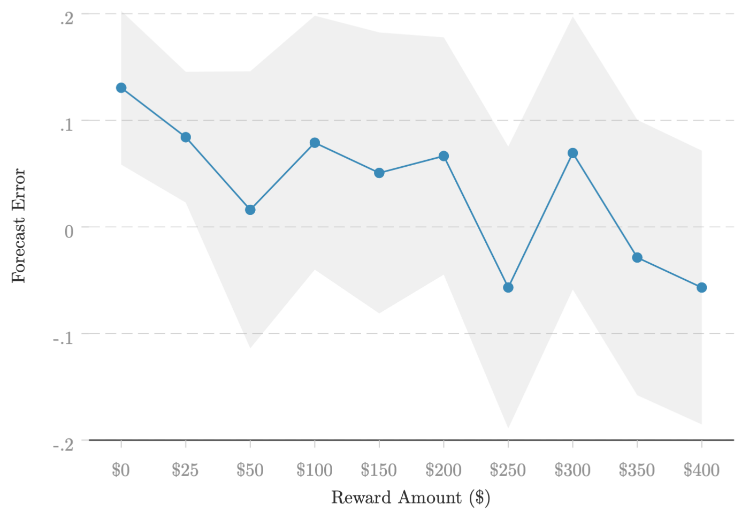
Note: Data for online firms comes from 5,800 survey responses on the Stanford-TechCo Study of Internet Entrepreneurship

Table B.1: Effect of the Reward treatment on Additional Question Timing

	Time (s) Past Sales	TimingGoodBad3Months	TimingProb
Reward '00s	-1.206	0.678***	0.306
	(1.443)	(0.255)	(0.234)
Dep. Mean	70.470	22.615	23.899
Observations	1167	3528	3525

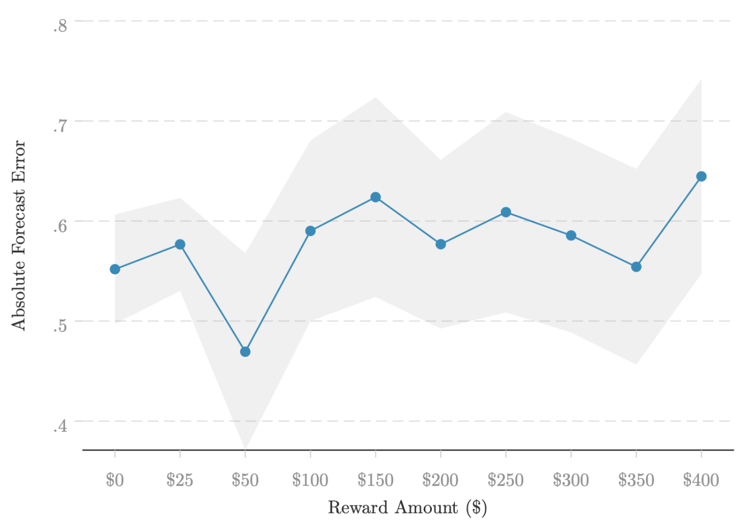
Note: Past Sales are asked about prior to the reward treatment, while question on good and bad cases and probabilities of outcomes occur after.

Figure B.2: Forecast Error vs. Reward Amount ($)



Note: Binscatter of log (forecast next quarter sales) – log (realization of next quarter sales) on the reward payment for forecasts within 10% of actual. Data from rounds 5 through 7, with standard errors clustered at the firm level

Figure B.3: Squared Forecast Error vs. Reward Amount ($)



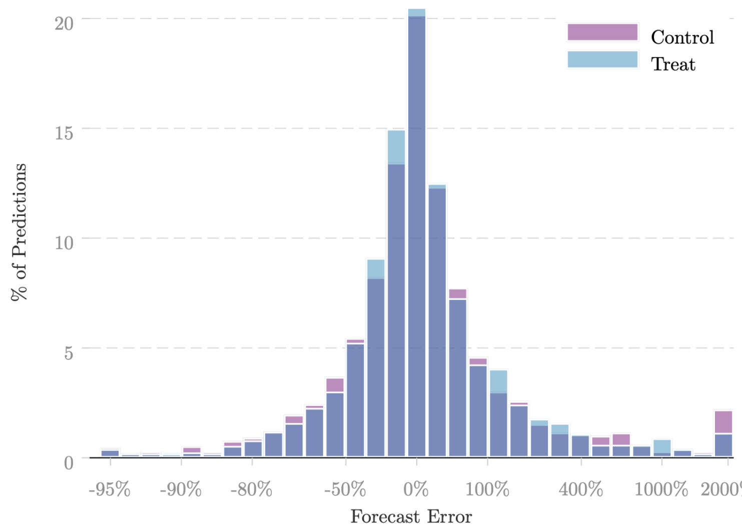
Note: Binscatter of squared value of log(forecast next quarter sales) – log(realization of next quarter sales) on the reward payment for forecasts within 10% of actual. Data from rounds 5 through 7, with standard errors clustered at the firm level.

C Dashboard

Figure C.1: Forecasting Performance By Dashboard Usage, baseline

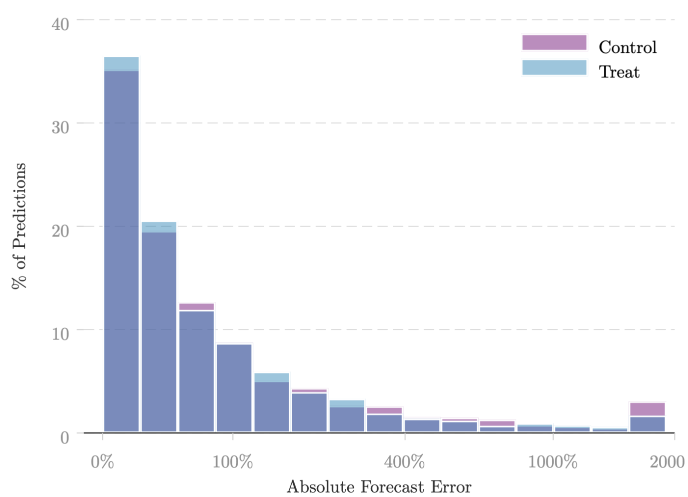
Notes: Data for firms comes from survey responses on the Stanford Study of Internet Entrepreneurship. Predictions were gathered in the Summer of 2019 (round 2), prior to the COVID-19 pandemic, in responses to the question "What do you predict your revenue on TechCo will be over the next there months?" Forecast error was then calculated by comparing their predictions with sales data recorded by TechCo. Dashboard usage was calculated based on the number of times they viewed their TechCo Account.

Figure C.2: Forecasting Error by Dashboard Treatment



Note: Forecasting error is calculated as log(forecast next quarter sales) – log(realization of next quarter sales). Data is from rounds 5 through 7.

Figure C.3: Forecasting Error Squared by Dashboard Treatment



Note: Squared Forecasting error is calculated as the square of log(forecast next quarter sales) – log(realization of next quarter sales). Data is from rounds 5 through 7.

Table C.1: Dashboard Treatment Effects by % Sales on TechCo

	Report Err.	(Report Err.)^2	Forecast Err.	(Forecast Err.)^2
Low Stripe User X Dashboard	-0.053*	-0.329***	-0.062	-0.206**
	(0.031)	(0.062)	(0.046)	(0.082)
High Stripe User X Dashboard	-0.001	-0.143***	0.022	-0.054
	(0.025)	(0.049)	(0.033)	(0.060)
F-Test Stripe Use	0.152	0.008	0.101	0.086
Time FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Dep. Mean	0.107	0.695	0.099	1.034
Observations	6813	6813	6658	6658

Regression of log of (forecast next quarter sales) – log (realization of next quarter sales) on the dashboard treatment for forecasts within 10% of actual. Data from rounds 2 through 7, with standard errors clustered at the firm level.

Table C.2: Dashboard Treatment Effects by Revenue

	Report Err.	(Report Err.)^2	Forecast Err.	(Forecast Err.)^2
Low Rev. X Dashboard	-0.065*	-0.354***	-0.055	-0.207**
	(0.035)	(0.072)	(0.050)	(0.096)
High Rev. X Dashboard	0.010	-0.083*	0.022	-0.045
	(0.024)	(0.045)	(0.030)	(0.051)
F-Test Revenue	0.054	0.001	0.151	0.104
Time FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Dep. Mean	0.107	0.695	0.099	1.034
Observations	6101	6101	6435	6435

Regression of log of (forecast next quarter sales) – log (realization of next quarter sales) on the dashboard treatment for forecasts within 10% of actual. Data from rounds 2 through 6, with standard errors clustered at the firm level.

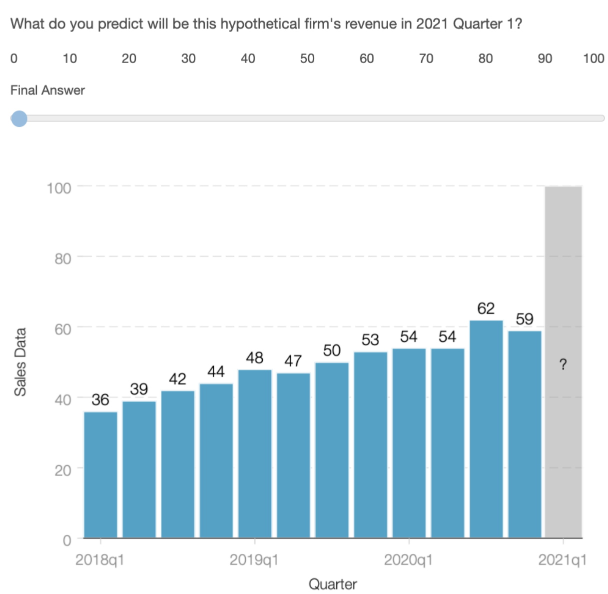
Table C.3: Dashboard Treatment Effects by Dashboard Usage

	Report Err.	(Report Err.)^2	Forecast Err.	(Forecast Err.)^2
Low Views X Dashboard	-0.056*	-0.292***	-0.038	-0.131
	(0.031)	(0.066)	(0.047)	(0.088)
High Views X Dashboard	0.003	-0.146***	0.008	-0.102*
	(0.027)	(0.048)	(0.032)	(0.056)
F-Test Views	0.116	0.055	0.377	0.763
Time FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Dep. Mean	0.107	0.695	0.099	1.034
Observations	6310	6310	6659	6659

Regression of log of (forecast next quarter sales) – log (realization of next quarter sales) on the dashboard treatment for forecasts within 10% of actual. Data from rounds 2 through 6, with standard errors clustered at the firm level.

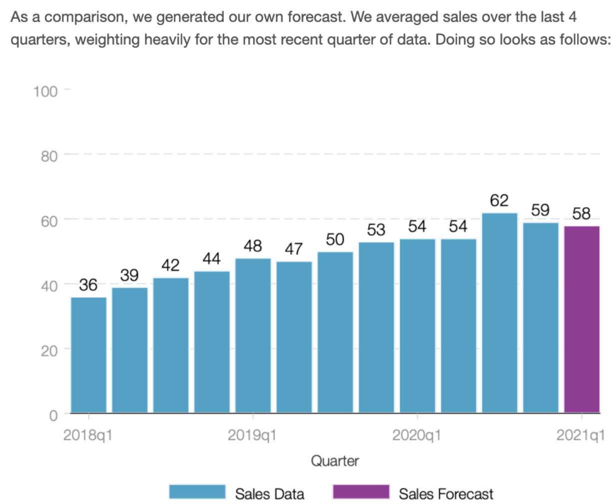
D Training

Figure D.1: Training Prompt



Note: Training prompt shown to respondents in Round 7 of the survey

Figure D.2: Training Reveal



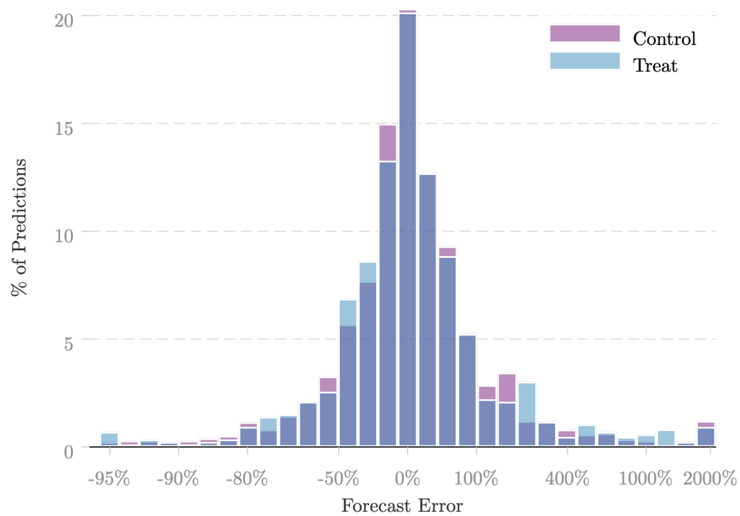
Note: Training reveal shown to respondents in Round 7 of the survey

Table D.1: Current Revenue Vs. Lag Revenue

	AsinhRev	AsinhRev	AsinhRev	AsinhRev
L1AsinhRev	0.997***	0.793***	0.740***	0.729***
	(0.000)	(0.003)	(0.004)	(0.005)
L2AsinhRev		0.204***	0.174***	0.153***
		(0.003)	(0.005)	(0.005)
L3AsinhRev			0.084***	0.037***
			(0.003)	(0.004)
L4AsinhRev				0.080***
				(0.003)
Dep. Mean	9.570	9.620	9.665	9.699
Coef. Sum	0.997	0.997	0.998	0.999
R-Squared	0.986	0.989	0.989	0.990
Adj R-Squared	0.986	0.989	0.989	0.990
# Obs	309896	274688	241655	210676

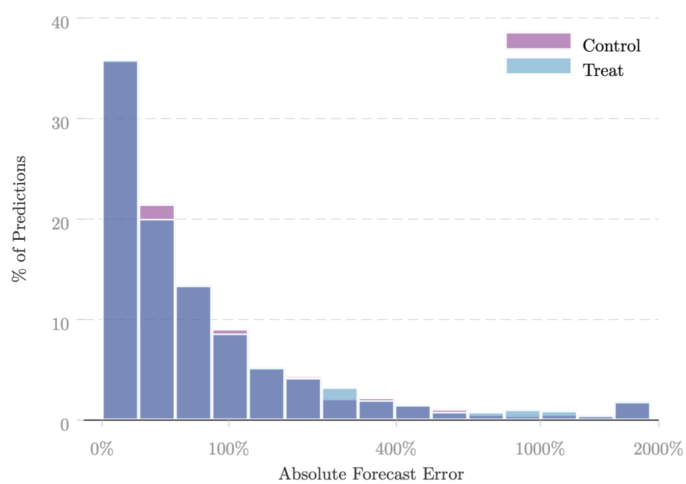
Note: Calculated using all 26,000 firms that were sampled prior to round 6

Figure D.3: Forecasting Error by Training Treatment



Note: Forecasting error is calculated as log(forecast next quarter sales) – log(realization of next quarter sales). Data is from rounds 5 through 7.

Figure D.4: Forecasting Error Squared by Training Treatment



Note: Squared Forecasting error is calculated as the square of log(forecast next quarter sales) – log(realization of next quarter sales). Data is from rounds 5 through 7.

E Management Questions

In the second round of the survey (Summer 2019), firms were asked to respond to a module on management practices. The questions were copied with minor adjustments from the Management and Organizational Practices Survey (MOPS) and the Annual Survey of Entrepreneurs (ASE).

- How many key performance indicators (KPIs) are monitored at your business?
- How frequently are KPIs typically reviewed at your business?
- What did you do when a service or production problem arises in your business?
- What describes the time frame of your service/production targets?
- How easy or difficult is it to achieve service, or production targets?
- What are the primary ways employees are promoted in your business?
- When is an under-performing employee reassigned or dismissed?