Corporate Prediction Markets: Evidence from Google, Ford, and Firm X[1]

September 2013

Bo Cowgill
Haas School of Business, UC Berkeley

Eric Zitzewitz
Dartmouth College and NBER

Despite the popularity of prediction markets among economists, businesses and policymakers have been slow to adopt them in decision making. Most studies of prediction markets outside the lab are from public markets with large trading populations. Corporate prediction markets face additional issues, such as thinness, weak incentives, limited entry and the potential for traders with ulterior motives – raising questions about how well these markets will perform. We examine data from prediction markets run by Google, Ford and Firm X (a large private materials company). Despite theoretically adverse conditions, we find these markets are relatively efficient, and improve upon the forecasts of experts at all three firms by as much as a 25% reduction in mean squared error. The most notable inefficiency is an optimism bias in the markets at Google and Ford. The inefficiencies that do exist generally become smaller over time. More experienced traders and those with higher past performance trade against the identified inefficiencies, suggesting that the markets' efficiency improves because traders gain experience and less skilled traders exit the market.

---

[1] This paper incorporates material on the efficiency of Google's prediction markets from a paper entitled "Using Prediction Markets to Track Information Flows: Evidence from Google;" that material will not be published separately. Justin Wolfers was a co-author of that earlier paper, but withdrew from this project due to conflicting obligations. The authors would like to thank Ford, Google, and Firm X for sharing the data used in this paper and Inkling Markets for facilitating the sharing of the Ford and Firm X data. We thank Susan Athey, Gary Becker, Jonathon Cummings, Stefano DellaVigna, Harrison Hong, Larry Katz, Steven Levitt, Ulrike Malmendier, John Morgan, Kevin M. Murphy, Michael Ostrovsky, Marco Ottaviani, Paul Oyer, Parag Pathak, Tanya Rosenblat, Richard Schmalensee, Jesse Shapiro, Kathryn Shaw, Noam Yuchtman and seminar participants at the AEA meetings, AMMA 2009, Berkeley, Chicago, Google, the Kaufmann Foundation, INFORMS, the NBER Summer Institute, the Stanford Institute for Theoretical Economics, and Wesleyan for helpful suggestions and comments. At Google, we thank Doug Banks, Patri Friedman, Ilya Kyrnos, Piaw Na, Jonathan Rosenberg and Hal Varian for helpful comments. At Ford we thank Michael Cavaretta and Tom Montgomery, and at Inkling Markets, we thank Pat Carolan and Adam Siegel. Cowgill was a full-time employee of Google from 2003 to 2009, and since 2009 he has received financial support and research data. Zitzewitz was a visiting scientist at Google for part of 2008. He was also an employee of Ford during the summers of 1988-1990.

Corporate Prediction Markets: Evidence from Google, Ford, and Firm X

The success of public prediction markets such as the Iowa Electronic Markets has led to considerable interest in running prediction markets inside organizations. Interest is motivated in part by the hope that prediction markets might help aggregate information that is trapped in hierarchies for political reasons, such as perceptions that messengers are punished for sharing bad news (e.g., Prendergast, 1993). A popular book arguing the benefits to organizations from harnessing *The Wisdom of Crowds* (Suroweicki, 2004) was a notable source of enthusiasm.

Markets in organizations face issues distinct from public prediction markets, however. If markets are run on topics of strategic importance, there is often a need to limit participation for confidentiality reasons. Limited participation makes markets thinner. In thinner markets, biases in participants' trading may have more influence on prices. Employees may optimistically bias their trading in order to influence management's view of their projects' performance or prospects. In addition to strategic biases, members of an organization may not be sufficiently dispassionate when making predictions. Employees may select employers based partly on optimism about their future, and belonging to an organization may likewise engender a favorable view of its prospects. Employees may suffer from other biases, such as probability misperceptions or loss aversion. Whereas in public prediction markets arbitrageurs may enter to eliminate any resulting inefficiencies, in corporate prediction markets, this entry may be less feasible.

This paper examines the efficiency of corporate prediction markets by studying markets at three major companies: Google, Ford Motor Company, and Firm X. These firms' markets were chosen because they are among the largest corporate markets we are aware of and they span the many diverse ways that other companies have employed prediction markets. Our sample includes all of the major types of corporate prediction markets we are aware of, including markets that forecast demand, product quality, deadlines being met, and external events. It includes both markets into which the entire company was invited to trade and markets available only to hand-picked employees or specific divisions. It also includes diversity in the strength of incentives and in market mechanisms and design. Table 1 summarizes these characteristics and shows examples of other major corporations that we are aware of having used markets similar to those in our sample.

Despite large differences in market design, operation, participation, and incentives, we find that prediction market prices at our three companies are well-calibrated to probabilities and improve upon alternative forecasting methods. Ford employs experts to forecast weekly vehicle sales, and we show that contemporaneous prediction market forecasts outperform the expert forecast, achieving a 25% lower mean squared error (p-value 0.104). Google and Firm X did not have formal expert forecasts of the variables being predicted by its markets, but for markets forecasting continuous variables, expert opinion was used in the construction of the securities. Google and Firm X created securities tracking the probability of the outcome falling into one of 3 or more bins, and an expert was asked to create bin boundaries that equalized ex ante probabilities. Firm X also ran binary markets on whether a variable would be above or below an "over/under" median forecast. At both Google and Firm X market-based forecasts outperform those used in designing the securities, using market prices from the first 24 hours of trading so that we are again comparing forecasts of roughly similar vintage.

The strong relative predictive performance of the Google and Ford markets is achieved despite several pricing inefficiencies. Google's markets exhibit an optimism bias and an overreaction to new information. Ford's Sales also markets exhibit an overreaction to new information, and its markets on the popularity of new car Features exhibit an optimism bias and an under reaction to new information. While the Features markets ended too soon for us to access efficiency over time, we find that the inefficiencies in the Google and Ford Sales markets disappear by the end of the sample. Improvement over time is driven by two mechanisms: first, more experienced traders trade against the identified inefficiencies and earn higher returns, suggesting that traders become better calibrated with experience. Second, traders (of a given experience level) with higher past returns earn higher future returns, trade against identified inefficiencies, and trade more in the future. These results together suggest that traders differ in their skill levels, that they learn about their ability over time, and that self-selection causes the average skill level in the market to rise over time.

Our Google data, which include information on traders' job and product assignments, allow us to examine the role played by insiders in corporate markets. If we define an insider narrowly, as a team member for a project that is the subject of a market, or as a friend of a team member (as reported on a social network survey), we find that insiders account for 10 percent of trades, that insiders are more likely to be on the optimistic side of a market, and that insiders' trades are not systematically profitable or unprofitable. If we instead define insiders more broadly, as those traders we would expect to be most central to social and professional networks at Google (software engineers located at the Mountain

View headquarters with longer tenure), we find that these traders are less optimistic and more profitable than other traders. So while a small number of insiders may trade optimistically in markets on their own projects, perhaps reflecting either overconfidence or ulterior motivations, they are offset by a larger group of traders who also have relevant expertise and fewer professional reasons to be biased.

Taken together, these results suggest that despite limited participation, individual traders' biases, and the potential for ulterior trading motives, corporate prediction markets perform reasonably well, and appear to do so for reasons anticipated by theory. Equilibrium market prices reflect an aggregation of the information and any subjective biases of their participants (Grossman, 1976; Grossman and Stiglitz, 1980; Ottaviani and Sorensen, 2007a). Traders with an outside interest in manipulating prices may attempt to do so (Allen and Gale, 1992; Aggarwal and Wu, 2006; Goldstein and Guembel, 2008), but, as emphasized by Hanson and Oprea (2009), the potential for manipulation creates incentives for other traders to become informed. Similar logic applies to traders with subjective biases -- their presence creates incentives for participation by informed traders. Our results of initial inefficiency disappearing, with more experienced and skilled traders trading against the inefficiencies, are consistent with this set of predictions.

Our paper contributes to an increasingly extensive empirical literature on prediction markets and a much smaller literature describing experimental markets run at companies. Forsythe, et. al. (1992) and Berg, et. al. (2008) analyze the empirical results from the Iowa Electronic Market on political outcomes, finding that markets outperform polls as predictors of future election results. Wolfers and Zitzewitz (2004) and Snowberg, Wolfers, and Zitzewitz (2005, 2012) examine a broader set of markets, again concluding that prediction markets at least weakly outperform alternative forecasts. A series of papers have used prices from public prediction markets to estimate the effects of policies and political outcomes (e.g., Rigobon and Sack, 2005; Snowberg, Wolfers and Zitzewitz, 2007a and 2007b; Wolfers and Zitzewitz, 2009).

While most of this literature is empirical, Ottoviani and Sorenson (2007b) present a theoretical framework for prediction markets inside organizations. The empirical literature begins with Ortner (1998), which reports on markets run at Siemens about project deadlines. Chen and Plott (2002) and Gillen, Plott, and Shum (2013) report on sales forecasting markets run inside Hewlett-Packard and Intel, respectively. Hankins and Lee (2011) describe three experimental prediction markets run at Nokia, including one predicting smart phone sales. Most of these experiments are much smaller than the markets we study. The largest is the sales forecasting experiment at Intel, which is of roughly comparable scale to the sales forecasting portion of the markets run at Ford.[2]

Our study differs from these prior and concurrent studies in several ways. First, the larger scale of the markets we analyze allows us to test for market inefficiencies with great statistical power, as well as to characterize differences in efficiency over time and across types of markets. Second, the microdata available on Google participants allow us to identify the characteristics of employees who trade with and against inefficiencies. Third, the markets we analyze are non-experimental in the sense that they were initiated by the companies themselves.[3] They are thus more field than field experiment. While a downside to field data is that some research opportunities may have been missed, an advantage is that the markets we study are more likely to be representative of prediction markets as companies will implement them in the future.

Prior research informs our analyses of the specific inefficiencies we examine. Building on Ali's (1977) analysis of horseracing parimutuel markets, Manski (2006) shows that two common features of prediction markets -- budget constraints and the skewed payoff structure of binary securities -- can combine to cause a longshot bias in which prices of low-priced securities will be upwardly biased relative to median beliefs. Gjerstad (2005), Wolfers and Zitzewitz (2005), and Ottaviani and Sorensen (2007a) generalize this result to a broader set of risk preferences and information environments, showing that the sign of any bias is ambiguous.

The optimistic bias we document could either arise from genuine optimism or a conscious effort to manipulate prices. As Hanson and Oprea (2009) argue, the extent to which a (consciously or unconsciously) biased trader will affect prices depends on the ability of other traders to become informed and enter the market. In past episodes of apparent price manipulation in public prediction markets, other traders entered and traded against the apparent manipulation, reducing its impact on prices.[4] The price impact of manipulators in experimental markets is examined by Hanson, Oprea, and Porter (2006) and Jian and Sami (2012), with the former concluding that manipulation does not affect the accuracy of prices and the latter concluding that effects depend on the correlation of signals given to

[2] The Intel sales forecasting markets cover 46 product*period combinations, while the sales forecasting component of Ford's markets cover 78 product*period combinations (6 models times 13 weeks).
[3] The markets at Google were created by a group that included an author on this paper (Cowgill), but several years prior to his beginning his career as an economist.
[4] See Wolfers and Zitzewitz (2004 and 2006), Rhode and Strumpf (2004 and 2008), Hansen, Schmidt, and Strobel (2004), and Newman (2012) for discussions.

participants. In the field, the robustness of a corporate prediction market may depend on the ability and willingness of unbiased traders to enter the market and become informed, which may be constrained by limited participation.

To the extent that the optimistic bias we document is behavioral, our results also speak to the growing literature about overconfidence and excess optimism in organizations. Recent work shows that worker overconfidence has significant economic consequences for workers and firms. A theoretical literature explores how optimism may improve motivation of employees (Benabou and Tirole, 2002 and 2003; Compte and Postlewaite, 2004), reduce compensation costs (Oyer and Schaefer, 2005; Bergman and Jenter, 2007) or lead to risk-taking that generates positive externalities (Bernardo and Welch, 2001; Goel and Thakor, 2008). Empirical work suggests that the benefits of optimism-induced risk taking may be mixed. Hoffman (2011) finds that overconfidence causes truckers to select more training, and Larkin and Leider (2012) finds it causes employees to select more convex incentive contracts. In both cases, employee overconfidence lowers costs for firms. In contrast, Malmendier and Tate (2008) find that overconfident CEOs undertake mergers that are associated with lower stock performance for their employers. Corporate prediction markets provide tools for both measuring and potentially correcting employee optimism. Indeed, Firm X told us that a primary motivation for running markets was a desire to help senior managers become better calibrated forecasters.

The remainder of the paper is organized as follows. The next section provides background on the markets at Google, Ford, and Firm X. The following section presents our empirical analysis of the efficiency and inefficiencies of these markets. A discussion concludes.

1. Background on the Corporate Prediction Markets

The three companies whose prediction markets we examine, Google, Ford, and Firm X, are in different industries, have distinct corporate cultures, and took different approaches in their prediction market implementations. We will describe them in turn, and then discuss commonalities and differences.

1.1 Background on the Companies and Their Markets

Google is a software company, headquartered in Mountain View, CA, with a highly educated workforce and a high level of internal transparency. Its prediction markets began as a "20% time project" initiated by a group of employees that included a co-author of this paper (Cowgill) prior to his PhD. Google opened its prediction markets to all employees.

The focus of Google's markets were whether specific quarterly "Objectives and Key Results" (OKRs) would be achieved. OKRs are goals of high importance to the company (e.g., the number of users, a third-party quality rating, or the on-time completion of key products). The attainment of OKRs was widely discussed within the company, as described by Levy (2011):

> OKRs became an essential component of Google culture. Four times per year, everything stopped at Google for division-wide meetings to assess OKR progress. [...]
>
> It was essential that OKRs be measurable. An employee couldn't say, ``I will make Gmail a success" but, ``I will launch Gmail in September and have a million users by November." ``It's not a key result unless it has a number," says [senior executive] Marissa Mayer.

Google's markets were run with twin goals: 1) aggregating information for management about the success of an important project and 2) further communicating management's interest in the success of the project. Prediction market prices were featured on the company intranet home page, and thus were of high visibility to employees. One particular anecdote illustrates how the markets impacted executive behavior. At a company-wide meeting, a senior executive made the following comment:

> [...] I'd like to talk about one of our key objectives for the last six quarters. During this entire time, one of our quarterly objectives has been to hire a new senior-level executive in charge of an important new objective to work on [redacted].
>
> We have failed to do this for the past six quarters. Judging from the [internal prediction markets], you saw this coming. The betting on this goal was extremely harsh. I am shocked and outraged by the lack of brown-nosing at this company [laughter].
>
> We've decided to look into the problem and figure it out, and I think we have gotten to the bottom of it. We've made some adjustments in the plans for the new team, and made some hard decisions about exactly what type of candidates we're looking for. … We're expecting to finally get it done in the upcoming quarter -- which would take this objective off the list once and for all.

The objective in question was indeed completed that quarter.

While the prediction market project aspired to cover every company-wide OKR, information on some projects needed to be too compartmentalized for them to be appropriate for a prediction market with mass participation. Thus a cost of wide participation was that some topics were necessarily off limits. Despite this, over 60 percent of quarterly OKRs were covered by markets.

The markets on OKRs spanned the topics typically covered in other corporate prediction markets, including demand forecasting, project completion, and product quality (Table 1). Demand forecasting markets typically involved an outcome captured by a continuous variable (e.g., "How many Gmail users will there be by the end of Q2?"). An expert was asked to partition the continuum of possible outcomes into five equally likely ranges. In contrast, project completion and product quality OKRs were more likely to have binary outcomes (e.g., would a project be completed by the announced deadline), and these markets had two outcome securities. In addition to markets on OKRs, Google also ran markets on other business-related external events (e.g., will Apple launch a computer based on Intel's Power PC chip) and on fun topics that were designed to increase participation in the other markets.

Ford is an automobile maker, headquartered in Dearborn, Michigan, with operations in many countries. Ford chose to focus its prediction markets on two topics of especially high importance: forecasting weekly sales volumes and predicting which car features would be popular with customers (as proxied in the interim by focus groups). Ford limited participation to employees with relevant expertise (in the marketing, engineering, and vehicle design departments).

Sales forecasting is an important activity at an automaker, as it is essential for planning procurement and production so as to minimize parts and vehicle inventories. Ford has a long history of employing experts to forecast sales and other macroeconomic variables. Sales forecasting is also a common application for prediction markets: some of the Google OKRs involved future use of its products, and sales forecasts were the subject of markets at Hewlett-Packard (Chen and Plott, 2002) and Intel (Gillen, Plott, and Shum, 2013). Like H-P and Intel, Ford has an expert make official sales forecasts, with which we can compare the contemporaneous forecast of the market for accuracy. Unlike in the Google markets, in the Ford sales forecasting markets, a single security was traded with a payoff that was a linear function of the weekly sales for a particular model.

The features markets run by Ford were markets that sought to predict the success of a decision prospectively, which are sometimes called decision markets (Hanson, 2002). Rather than defining success as a feature's long-term success in the marketplace, Ford chose feedback from focus groups as a more immediate outcome measure. A market asked whether a series of potential car features (e.g., an in-car vacuum) would reach a threshold level of interest in a focus group at a specified price. Each feature was represented by a binary security that paid off if the threshold was reached.

In the decision markets that have been discussed elsewhere in the literature (e.g., Hanson 2002; Wolfers and Zitzewitz, 2006), trades are unwound if the decision is not taken, so the price of a security reflect an expected outcome conditional on the decision. Ford's decision markets differed in that the features were always shown to focus groups eventually, so trades were never unwound. Rather than being used to decide whether to show a feature to focus groups, the market's expectation of the focus group outcome was combined with the actual outcome to provide a less noisy assessment. Focus groups are expensive to run, sample sizes are necessarily small, and so sampling errors can be meaningful. In contrast, opinions of employees may be cheaper to obtain, but employees are potentially biased, which is the reason focus groups are used in the first place. By asking employees to predict the results of focus groups, Ford sought to increase sample sizes while mitigating any biases in employee opinion.[5] In a 2011 press release, Ford mentioned that it decided against including a Ford-branded bike carrier and an in-car vacuum in future models based on trading in its Features prediction market.[6]

Firm X is a diversified basic materials firm headquartered in the United States. It refines crude oil, transports oil and petroleum products, and manufactures products including chemicals, building materials, paper products, and synthetic fibers like spandex. Many of its businesses are very sensitive to the macro-economy and/or to commodity prices, both of which were quite volatile during the period their markets ran (March 2008 to present). Firm X decided to focus its prediction markets on macroeconomic and commodity prices that were relevant to its business. Some of these variables were already priced by existing futures markets (e.g., the future level of the Dow Industrials index or the West Texas Intermediate crude oil price) and some are the subject of macroeconomic forecasting (e.g., the

[5] After the feature markets we analyze, Ford ran subsequent markets in which security payoffs were based on final markets prices, rather than focus group outcomes. These markets were beauty contests, with multiple equilibria. While the relative predictive performance of these markets would be an interesting research question, we were unable to obtain access to the focus group outcome data that corresponded to these markets. We therefore exclude these markets from our analysis.
[6] See http://www.internetnews.com/bus-news/article.php/3925571/Ford+Taps+CloudBased+Prediction+Market.htm (Last accessed 9/30/2013).

unemployment rate and general price inflation), but many others were not (e.g., the Spandex price in China, the Kansas City Fed's Financial Stress Index). In addition, markets were run on policy and political outcomes of interest to Firm X, such as bailouts, health care reform, and the midterm and Presidential elections.

Firm X's markets were started by a Senior Manager in Firm X's strategic planning department, and participation was limited to a hand-selected group of employees with relevant expertise. While the number of participants in the Firm X markets was much smaller than at Google or Ford, 57 out of 58 invitees participated, and the average participant placed 220 trades (compared with 48 at Google and 10 at Ford).

Firm X's market creator had an additional motivation beyond obtaining forecasts. "People are overconfident in their predictions," he says. "They either say 'X will happen' or 'X won't happen.' They fail to think probabilistically, or confront their mistakes when they happen. The market therefore changes the way participants think, and I believe this not only improves our forecasts but has a positive spillover on everything else our team does." This stated goal is particularly interesting in light of our results, which suggest that markets are initially optimistic and overconfident (e.g., they display a bias away from a naive prior), but that these biases decline over time and that more experienced traders trade against them.[7]

Just under 60 percent of Firm X's markets predicted a continuous variable. About one-fifth of these markets divided the continuum of possible future outcomes into 3-10 bins as in Google's markets, while almost all of the other 80 percent specified a single "over/under" threshold. A very small number of markets (18 out of 1345) used the linear payoffs used by Ford's Sales markets. For the remaining 40 percent of markets that predicted a discrete event (e.g., would President Obama be re-elected), there was a single security, which paid off if the specified event occurred.

1.2 Commonalities and Differences

Table 1 summarizes the types of markets run by the three companies, and provides examples of a few other companies we are aware of that have run related markets. All six types of markets we are aware of being run at other firms were run at our three firms. Google ran markets of all varieties, while Ford

[7] Although ironically, it was the markets at Google and Ford that displayed evidence of inefficiencies that disappeared over time; our analysis suggests that the Firm X markets were well-calibrated from the beginning.

focused on sales forecasting and decision markets, and Firm X's focused on external events. A few other firms have run many types of prediction markets (e.g., Eli Lilly, Best Buy), while others have run more focused experiments with one particular type of market.[8]

Table 2 contrasts the scale and some key features of our three markets. One important difference was the structure of the securities in the markets. As discussed above, Google used multiple bins for continuous outcomes (e.g., demand) and two bins for discrete outcomes (e.g., deadlines). In contrast, Ford used securities with linear payoffs for continuous outcomes and single binary securities for discrete outcomes. With a very small number of exceptions, Firm X used single binary securities for discrete outcomes and either bins or a single binary security combined with an "over/under" threshold for continuous outcomes.

The choice between two bins and single binary securities for discrete outcomes can potentially affect market efficiency if some participants exhibit "short aversion" (i.e., prefer to take positions by buying rather than selling). With bins, choices of boundaries can affect efficiency if participants take cues from them, as the literature on partition dependence suggests some do (Fox and Clemen, 2005; Sonnenman, et. al., 2011). We will test whether pricing suggests bias towards buying, as well as whether there is a bias towards pricing each of N bins at 1/N.

Two other important differences were the market making mechanism and the incentives provided to participants, which we discuss in turn.

1.2.1 Market-making Mechanism

Google used an approach similar to the Iowa Electronic Markets (see, e.g., Forsythe, et. al., 1992), in which the range of possible future outcomes is divided into a set of mutually exclusive and completely exhaustive bins, and securities are offered for each. For continuous variable outcomes, such as future demand for a product, five bins were typically used, with the boundaries chosen by an expert to roughly equalize ex ante probability. For OKRs with discrete outcomes, like whether a deadline or quality target will be met, there are generally two outcomes, and no reason to expect the *ex ante*

---

[8] We base these statements on public comments made at conferences by firms, as well as on interviews. In the latter case, we do not identify specific firms (e.g., the reference to "other pharma") unless we have received permission to, and we omit some examples we are aware of for brevity. It is of course possible that firms have run markets we are unaware of.

probability to be 0.5 (indeed, Google's official advice on forming OKRs is that they should be targets that will be met 65% of the time).

As on the Iowa Markets, participants can exchange a unit of artificial currency for a complete set of securities or vice versa. This approach should limit the impact of any bias towards buying on prices, since participants can buy any of the possible outcomes, and other participants can simultaneously sell all outcomes if their bid prices sum to more than one. At the same time, Google did not have an automated market maker, although traders were observed placing arbitrage trades (e.g., selling all possible outcomes when their bid prices summed to greater than one or, more rarely, buying when their ask prices summed to less than one).

Ford and Firm X used prediction market software developed by Inkling Markets (http://www.inklingmarkets.com/). Inkling's software uses an automated market maker that follows the logarithmic market scoring rule described in Hanson (2003). The market maker allows trading of infinitesimal amounts at zero transaction costs, and moves its price up or down in response to net buying or selling. The automated market maker ensures that traders can always place trades, which helps avoid frustration and is particularly important when participation is limited. In cases where securities are linked to a mutually exclusive and exhaustive set of outcomes, the automated market maker ensures that their prices always sum to one. An issue with an automated market maker is that it must be set at an initial price, and market prices can therefore be biased towards this initial bias, especially if participation is limited. Furthermore, if the initial price differs from a reasonable prior, then easy returns can be earned by being the first to trade. If relative performance (e.g., "bragging rights") is a source of motivation for trades, having performance depend to heavily on simply being the first to trade against an obviously incorrect price can be counterproductive. As a result, Inkling users take some care in setting initial prices, or in setting bin boundaries so that initial prices of 1/N are appropriate.

Thus the use of Inkling mechanism could potentially reinforce potential biases toward pricing at 1/N discussed above. We will test whether prices at Ford and Firm X are biased towards their initial starting values, particularly early in markets' life, when compared with the markets at Google.

1.2.2. Incentives

Modest incentives for successful trading were provided at all three firms. Monetary incentives were largest at Google, although even these were quite modest. Google endowed traders with equal amounts of an artificial currency at the beginning of each quarter, and at the end of each quarter, this currency was converted linearly into raffle tickets for traders who placed at least one trade. The prize budget was $10,000 each quarter, or about $25-100 per active trader. The raffle approach creates the possibility that a poorly performing trader may win a prize through chance, but has the advantage of making incentives for traders linear in artificial currency. Awarding a prize to the trader with the most currency would create convex incentives, which could make low priced binary securities excessively attractive, potentially distorting prices.

Ford also used a lottery that created incentives that were linear in the currency used by the marketplace. For legal and regulatory reasons, it was not able to offer prizes to participants based outside the U.S., but we are told that these were a small share of participants in the markets we analyze.[9] Ford's incentives in North America were smaller than Google's, consisting of several $100 gift certificates.

Firm X did not offer monetary incentives for its traders, but publicized the most successful traders. The high participation rate of eligible Firm X traders suggests that the prediction markets were emphasized by management, and thus reputational incentives to perform should have been meaningful. If more attention was paid to the best performers than to the worst, the reputational incentives could have been convex in performance, encouraging risk taking. In particular, traders may have preferred the positively skewed payoffs of low-priced binary securities, potentially causing these securities to be mispriced.

Google also published league tables of the best performing traders, but any convexity may have been muted by the linear monetary incentives that were also provided. We therefore might expect low-priced binary securities to be more overpriced at Firm X than at Google. With smaller linear incentives for most of its participants, Ford might be expected to be an intermediate case between Google and Firm X.

2. Results

This section presents statistical tests in three subsections. The first examines whether forecasts from prediction markets improve on contemporaneous expert forecasts. The second tests for the market inefficiencies discussed above: optimism, a bias towards initial prices or 1/N, and a mispricing of small probabilities. This subsection also tests whether examine whether biases increase or decrease with time. The third subsection examines how trader skill and experience are related to future profitability

[9] We have thus far been unable to obtain a precise percentage.

and to whether one trades with or against the aforementioned biases. This subsection also uses data on job and project assignments at Google to examine how "insiders" trade in markets.

2.1. Markets versus Experts

Given that firms run prediction markets at least partly to obtain predictions, a natural research question is whether the predictions from markets outperform alternatives, including forecasts by expert forecasters or managers. We compare markets' predictions with three types of alternative forecasts. The first is a formal forecast from an expert forecaster. Ford forecasts weekly auto sales for different models, and for the six models covered by prediction markets, we can compare the expert's forecast with the prediction market forecast from immediately before the forecast was issued.[10]

A second type of forecast we compare with are percentile forecasts derived from bin boundaries used in constructing the prediction market securities. As mentioned above, in order to avoid minimize biases from either partition dependence effects or the initialization of market maker prices at 1/N, both Google and Firm X sought expert help in choosing bin boundaries to equalize ex ante probabilities. For example, at Google, the prediction market organizers would ask the Product Manager for the relevant product (e.g., the Gmail Product Manager for markets on new Gmail users) for assistance in creating the bins. These experts were encouraged to use whatever sources they desired to set these boundaries, and they often consulted historical data or made statistical forecasts. The bin boundaries they chose can be interpreted as specific-percentile forecasts, and it is straightforward to obtain an approximate median forecast from these boundaries.[11]

A third, related, type of forecast can be obtained from "over/under" markets that were run by Firm X on continuous variables. In these markets, a single security was traded that paid off if a macroeconomic variable exceeded a threshold, and as above that was chosen to create a 50 percent ex ante probability. The threshold can therefore be interpreted as a median forecast. About half of the binary markets in our sample used a prior-period value as the threshold (e.g., "will housing starts be up from last month?"). We analyze the other markets, in order to focus on instances where an over/under value was actively selected.

[10] The expert forecasts were issued 11 days before the week in question began. The six forecasted models were the Escape, F-150, Focus, Fusion, Super Duty, and Lincoln (all models). The official sales forecasts are closely held at Ford and were not available to the vast majority of predict market participants.

[11] For example, when there are an even number of bins, the boundary between the two middle bins is a median forecast. When there are an odd number of bins, the midpoint of the middle bin is an approximate median forecast.

Table 3 presents the results of these comparisons. In each column we report the results of horserace regressions (Fair and Shiller, 1989) of the security payoffs on the prediction market and expert forecasts. We also report the ratio of the prediction market and expert mean squared errors, and the p-value from a f-test for the equivalence of the two variances. In all four cases, the prediction market forecast has a lower mean squared error and receives a higher weight in the horserace regression.

The expert forecasts we study obviously differ in their formality. Ford has a long history of producing forecasts of weekly auto sales, which are clearly of high importance to planning procurement and production so as to minimize part and vehicle inventories. While the individuals setting the bin boundaries at Google and Firm X were chosen to be the most knowledgeable at the company, it is possible that less effort was put into their forecasts than was exerted at Ford. Nevertheless, it is interesting to note that the mean squared error improvement achieved by the prediction market at Ford is among the largest.[12]

2.2. Calibration and Inefficiencies

In this subsection, we test whether the markets at Google, Ford, and Firm X made yielded efficient forecasts, in the sense that the forecasts did not make errors that were predictable at the time of the forecast. This is equivalent to asking whether the markets yielded predictable returns. In particular, if a market is asked to forecast Y (which could be a binary variable indicating whether an event occurred, or a continuous variable indicating, e.g., the sales of a car model), then an efficient forecast at time t will be E(Y | Ht), where Ht is the set of information known publicly at time t. If prediction market prices are efficient forecasts, then the price at time t is equal to this expectation, Pt = E(Y | Ht), and expected future returns are zero, E(Y – Pt | Ht) = 0.

We focus our tests on variables that are known at time t and that our above review of the theory literature suggests may be correlated with mispricings. In particular, we ask whether future prediction market returns are correlated with the current price level, the difference between the current price and either the market maker's initial price or 1/N (where N is the number of mutually exclusive outcomes), and the extent to which a security is tied to an outcome that is good for the

[12] The p-value for the test for the statistical significance of the improvement is largest at Ford, at 0.104, but this is related to the much smaller sample size at Ford.

company. We also ask whether extreme outcomes are over or underpriced and whether the average return is systematically positive or negative, due to investors preferring to buy rather than sell.

Figures 1, 2, and 3 graph the future value of securities, conditional on current price for binary securities at Google, Firm X, and Ford, respectively. The prices and future values of binary securities range from 0 to 1, and trades are divided into 20 bins (0-0.05, 0.05-0.1, etc.) based on their trade price. The average trade price and ultimate payoffs for each bin are graphed on the x and y-axes, respectively.[13] A 95% confidence interval for the average payoff is also graphed, along with a 45-degree line for comparison. The standard errors used to construct the confidence interval are heteroskedasticity-robust and allow for two-dimensional clustering within markets and calendar months (using the procedure in Petersen, 2009).[14] Observations are weighted by time-to-next trade, which weights trades according to the amount of time that they persist as the last trade, and thus according to the likelihood they would be taken to be the current market forecast by a user consulting the market at a random time.[15]

Google and Firm X's markets appear approximately well-calibrated. Both markets exhibit an apparent underpricing of securities with prices below 0.2, and an overpricing for securities above that price level, but this is slight, especially for Firm X. In contrast, the Ford Feature markets appear to exhibit substantial overpricing for securities trading below 0.7. The initial price in Ford's Feature markets was always 0.5, and we initially thought that overpricing for securities trading near 0.5 could be related to a market-maker-induced bias towards the initial price due to thin trading that was more pronounced for unattractive features. Yet, a very similar pattern is observed if we restrict Figure 3 to trades after the 25th, 50th, or 100th in each security, or to those with significant numbers of both buys and sells. For example, Figure 4 sorts trades after the 25th for each security in Ford's features markets into bins based on the share of prior trades that were buys. It appears that Ford's markets overprice

---

[13] All securities in the Google markets and Ford feature markets were binary, none of the contracts in the Ford sales markets were (they had payoffs linear in vehicle sales). Almost all Firm X markets were binary – the exceptions were a small number of markets with linear payoffs in commodity prices. These markets accounted just under 1 percent of markets and trades, and they are excluded from Figure 3.

[14] Allowing for clustering at the market level allows for arbitrary correlations within the returns-to-expiry for trades within the same market: in this case for the fact that returns within securities will be positively correlated and returns across securities within markets will be negatively correlated. The Ford prediction markets were short-lived enough that we do not have a sufficient number of calendar months for clustering to be valid, so we instead use one dimensional clustering on markets (which, in the Sales markets, is also equivalent to clustering on time periods).

[15] Note that weighting in this manner does not produce a look-ahead bias from a forecasting perspective. Equal weighting trades produce very similar, albeit slightly noisier, results.

features that have mixed support among its employees, but that markets on the most popular features are more efficiently priced. In particular, the latter markets appear to take into account the fact that even the most popular features ex ante will not necessarily be well received with certainty by a given focus group. In contrast, the markets appear to not anticipate the fact that features receiving mixed support among employees appear to never obtain the requisite level of support in focus groups.

Figure 5 examines the calibration of Ford's sales markets. Given that these are linear markets and that they track sales for different models with differing overall sales levels, we scale prices and payoffs using a model's past sales. In order to ensure that we do not condition our analysis on information that market participants would not have observed, we use 3-week lagged sales. The x-axis plots the log difference between the sales forecast by a trade and lagged sales, and the y-axis plots the average difference between actual log weekly sales and lagged sales. The graph suggests that in contrast to the Features markets, the Sales markets are generally well-calibrated, albeit perhaps with a mild optimistic bias.

Table 4 presents regressions that test for predictabilities in returns, including those suggested by the figures. An observation in each regression is a trade, and the dependent variable is returns to expiry in percentage points (i.e., expiry value – price). The regressions include a constant, which captures any general over or underpricing, a variable capturing the difference between the trade price and a naïve prior, a variable capturing the "optimism" of a security (i.e., whether it tracked an outcome that was good for the company), and a variable capturing the extremeness of the outcome tracked by the security. The optimism variable is scaled from +1 for the best outcome to -1 for the worst.[16] The extremeness of a security is defined as the absolute value of optimism. For binary markets, every outcome is equally extreme, so the extremeness coefficient is identified in markets with 3 or more bins.

Two features of Ford's markets limit the analyses we can conduct. First, trading optimistically (i.e., predicting high sales or that a feature would be well received) always involved buying in their markets, so an optimism bias and a bias towards buying cannot be separately identified. This is not true at Firm X and Google, where these two biases can be separated. Second, since Ford's markets were

[16] In general, classifying outcomes as good or bad for the company involved few judgment calls. High demand or sales, high prices for products, high product quality, positive focus group responses, and projects completed on time were all judged to be positive outcomes. For fun markets, and for the few cases where we thought there was room for disagreement (e.g., is an Intel-based Mac good for Google? Are banking or auto bailouts good for Firm X Industries?), we left optimism uncoded.

either binary (the Features markets) or linear (the Sales markets), we cannot test whether extreme outcomes were over or underpriced.

Google's markets exhibit each of the four biases we test for. The negative constant implies that the average trade price is above its ultimate payoff. This is consistent with a preference for taking positions by buying rather than by exchanging currency for a complete set of securities and then selling. Controlling for this, securities with prices that are above a naïve prior (1/N, where N is the number of outcomes/bins) are overpriced. This is the opposite of the longshot bias predicted by the Ali(1977) and Manski (2006) models and is also inconsistent with participants taking cues from security boundaries as in the partition dependence literature. It is instead consistent with investors collectively underreacting to the information used in designing the boundaries or overreacting to other, potentially newer information. Google's markets also exhibit an overpricing of optimistic securities, and a slight underpricing of extreme outcomes.

In contrast to Google, Firm X's markets exhibit almost no evidence of bias. This is not simply due to imprecision of the estimates, as the Firm X sample is larger in terms of markets and securities, and coefficients of the magnitude found at Google can be rejected for the (Price – Prior) and optimism variables.

Our analyses for Ford are limited by both sample size and the fact that all of Ford's markets involved participants buying to express optimism and selling to express pessimism. For the Features markets, we find clear evidence of an optimism bias (as evidence by the negative average return), of a large bias in pricing towards the naïve prior, and of an overpricing of securities trading close to the prior. These biases are all apparent in Figure 3. The Sales markets appear more efficient, although the coefficients on the Price – Prior variables are large in magnitude and imprecisely estimated, so we cannot reject economically meaningful inefficiencies.

As discussed above, the Ford and Firm X prediction markets use an automated market maker that is initialized at prior value, and we might expect prices to be biased towards that initialization value, at least early in the life of the market. To investigate this possibility, we number the trades in each security sequentially and then split the sample according to this trade number (Table 5).[17] In the Ford Feature markets (Panel D), it does appear that trades before the 25th are dramatically biased toward the

[17] We take this approach to splitting the trading history of markets because the trade number is a variable that will be known at the time of the trade, while the whether a trade is in a given decile would not be known.

initialization value, as evidence by the positive coefficient on (Price – Prior). This bias remains large and ceases to decline after the 25th trade, however, and as mentioned above, it is present even for securities where there is both buying and selling activity, so it does not appear entirely related to biases induced by the market maker.

Furthermore, the Ford Sales market and Firm X markets use the same automated market maker as the Ford Feature markets, but do not exhibit similar biases, even very early in the life of these markets. Partly this may reflect effort having been put in to ensure that the initialization values were reasonable, so that even the first trade could move the market to a price that was no longer biased toward the initialization value. The large bias away from the prior in Ford Sales markets after trade number 50 turns out to be driven by a single, very inaccurate, market for one model in the first week; if that market is excluded, the coefficient on Price – Prior is consistent with other subsamples.

The Google markets did not use an automated market maker, and thus they have less reason to be biased towards the prior value early in their life. Indeed, the results in Table 5 imply that they are actually biased away from the prior early in their life and that this bias abates with more trading history. In contrast, the optimistic bias discussed above is small early in a market's life, and grows over time.

Table 6 disaggregates the Google and Firm X markets by topic and examines efficiency for each. The predictabilities in the Google markets appear to varying extents in each subset. The optimism bias arises from markets on whether projects would be completed as scheduled – in other words from markets on outcomes that are arguably most under Google's employees control, while markets on product quality or demand for Google's products (e.g., Gmail users) do not exhibit statistically significant optimistic biases and those on external news exhibit the opposite result.

The Firm X markets deal entirely with external events. Even when divided into ten subsamples, the Firm X markets exhibit little evidence of inefficiency. There are positive and significant coefficients on the (Price – Prior) variable for markets on commodity prices, energy, and Eurozone political events, implying that markets were biased towards their initial values in these areas. Markets on inflation exhibit an optimistic bias. Given the macro environment of the 2008-13 period and the fact that Firm X produces basic materials high inflation and high commodity and energy prices were coded as good for Firm X, so an optimistic bias means that Firm X's traders expected too much inflation. Markets on employment outcomes exhibited a pessimistic bias, implying that Firm X's traders expected too much unemployment. Given Firm X's affiliation with the Republican party, in politics markets Republican victories were coded as good for the firm, and thus the pessimistic bias in these markets implies that traders expected too little political success for Republicans (unsurprisingly, this result is driven by the 2010 election cycle). We do not discern a consistent theme in these results. We conclude that the Firm X's markets appear to have been efficient, with only slightly more statistically significant correlations than would have arisen through chance.

The fact that the optimism bias in Google's markets is largest in markets with outcomes that are under the control of Google employees is suggestive of a strategic bias, such as employees betting that projects will be completed on time in order to influence management's perception of their performance. As discussed above though, optimism could also arise for behavioral reasons as well. To investigate this possibility, we conduct tests for company-wide "mood swings" in prediction market pricing. In Cowgill and Zitzewitz (2013), we find daily frequency correlations between the company stock price and job satisfaction, physical output, hours worked, hiring decisions, and the evaluation of candidates and ideas. There is no persistence in these correlations (i.e., the stock price change from last week is not correlated with the outcome variables) which is inconsistent with standard explanations, such as an increase in employee wealth affecting labor supply decisions, or good news for a company affecting future labor demand and thus hiring. Instead we conclude that companywide "mood swings" are the likely explanation.

Table 7 presents tests for mood-swing effects on the size of the optimism bias at Google. The regressions repeat the specification in Table 4, with the optimism variable interacted with Google stock returns on days t+1, t, t-1, and t-2. In a variety of different specifications, we find that a 2% increase in Google's stock price (roughly a one standard deviation change) is associated with prediction market prices for securities tracking optimism outcomes being priced 3-4 percentage points higher, relative to their pricing on an average day. As in Cowgill and Zitzewitz (2013), these effects are quite temporary, as there is no association between the prediction market prices and day t-2 returns, as we would expect if the aforementioned relationship was driven by good news leading to both higher stock and prediction market prices.[18]

Finally, Table 8 presents tests of how the aforementioned (Price – Prior) and optimism biases evolved over our sample. Regressions from Table 4 are modified by the inclusion of a time trend (which is scaled to equal 0 at the beginning of the sample and 1 at the end) and interactions of the time trend with the bias variables. The results suggest that biases away from the prior in the Google and Ford Sales

[18] Unfortunately, we cannot conduct a similar analysis for the other market in our sample that exhibited an optimistic bias (the Ford Features market), given that most of the trading was on a limited number of days.

markets are largest at the beginning of the sample and essentially disappear by the end of the sample. The same appears to be true of the optimism bias in Google's markets. The analysis of Firm X's markets also reveals a short aversion bias at the beginning of the sample that is more than completely reversed by the end, although this result is not present in the subsample for which optimism can be signed, and thus appears to be driven by the fun markets. Unfortunately, the markets that appear most inefficient, the Ford Feature markets, changed to a beauty contest format after 2 weeks, and so an analysis of the evolution of the efficiency of these markets is not feasible.

2.3. Individual trader characteristics and market efficiency

This subsection analyzes which traders contribute to the biases discussed above, which traders trade against these biases, and which traders earn positive returns. For all three firms we have trader identifiers, and so we can construct variables that describe a trader's past history. For Google we also have data on traders' job and project assignments, and so we also construct variables that capture a trader's relationship with the subject of the market being traded.

In order to understand which traders contribute to and trade against pricing biases, we need analyze the relationship between the nature of a position being taken (e.g., its optimism) and the characteristics of the trader. We begin by analyzing all three companies, and thus focus on traders' past experience and past success. In the Google data, participants trade against each other, and thus every trade has a buyer and a seller. For Google, we structure the data so that each trade appears in the data twice (i.e., as a buy and as a sell). The characteristics of the security traded are first multiplied by the direction of that side of the trade (+1 if a buy, -1 if a sell) and they then regressed on trade fixed effects and the trader characteristics for that trade*side. This yields coefficients that are identical to what we obtain if we regress the security's characteristics on the difference in the characteristics of the buyer and seller, but has the advantage of facilitating the adjustment of standard errors for clustering within traders as well as within markets.[19] In the Ford and Firm X data, where participants trade against an automated market maker, we multiply the security characteristics by the direction of the trade (i.e., +1 if the participant is buying, -1 if selling) and then regress these on trader characteristics.

Table 9 presents the results of these tests. In Panel A, we find that Google traders with high past returns trade in a pessimistic direction, are more likely to sell than buy, and trade against securities

[19] Note that clustering by market also adjusts standard errors for the inclusion of two observations per trade, as clustering allows for any correlation of errors within cluster groups.

that are priced above 1/N. All three correlations are consistent with what the previous section found to be profitable, and consistent with this, we find that traders with high past returns earn high future returns. We also find that more experienced traders are more likely to sell and to trade against securities that are priced above 1/N, again in both cases consistent with what would be profitable. Thus we can conclude that less experienced traders and traders with less past success trade in a direction that would contribute to the biases discussed above.

In the Ford Sales markets, we also find that traders with more past experience and more past success are more likely to sell than buy (which means they are also trading pessimistically) and are more likely to sell when price is above its initial value (Panel B). The results presented in Section 2.2 suggest that trading in this direction should be profitable, and indeed we find a positive and significant relationship between future returns and both past performance and past experience.

Given that the Firm X markets did not display pricing biases, there is less reason to expect proxies for trader skill to be correlated with trading in a particular direction. Indeed, in Panel C, we see much less evidence of such correlations. We do see a positive correlation between past and future returns, consistent with traders displaying persistent skill.

Table 10 analyses the correlation between traders' future participation and their past performance. We sacrifice some comparability across the three markets in order to choose the optimal frequency for each. At Google, all traders are settled, returns are tallied, and prizes awarded at the end of each quarter, and participation in a subsequent quarter is done in a new account, so the quarter is the natural period of analysis. Our sample of Ford's sales markets only lasts for 13 weeks, and markets are resolved and returns realized each week, so the week is a more natural frequency for Ford. At Firm X, most markets were resolved at the monthly frequency (e.g., markets on monthly economic numbers), and so months are the natural time frequency at which investors would observe returns. We chose a future period far enough into the future that prior results would have been observed in time to inform the future participation decision (Quarter + 1 for Google, Week + 3 for Ford, Month + 2 for Firm X).

We find strong evidence that future participation is more likely for traders with strong past performance in the Google and Ford Sales markets, even when controlling for past trading activity levels. In unreported results, we find that these results are robust to many variations in the frequency of the analysis, length of the future lead, and the functional form of the independent variables. At Firm X, attrition from the prediction market is extremely rare, and does not appear related to past performance.

We turn to an analysis of correlations between prediction market trading and trader characteristics that are only available to us in the Google data. Table 11 presents regressions with the same structure as in Table 9, Panel A. We find that optimistic trades are made disproportionately by traders who are staffed on the project in question and by friends of those insiders (as indicated by either party on a social network survey). Insiders are also more likely to buy securities and to buy when securities are trading above 1/N. Consistent with this, they earn lower returns. Programmers and employees based in Mountain View and New York (Google's second largest office at the time of the study), who we might to be more knowledgable, tend to trade against biases and earn higher returns. The results are consistent with those with the most knowledge of a market's subject trading in an unprofitable (and potentially strategically biased) way, but with other knowledgable employees trading in the opposite direction, pushing prices back to their efficient level.

It is also interesting that while newly hired employees are more likely to sell than buy, they do trade more optimistically. It is worth noting that during this time period, the vast majority of new Google hires were hired directly from degree programs, and thus were inexperienced both in working at Google and in working in general. Therefore it is possible that their optimism reflected an initial miscalibration about the extent to which demand forecasts and deadlines are goals rather than forecasts. Consistent with this, we find that the correlation between hire date and optimism is strongest for markets on demand forecasts and on whether deadlines will be met.

3. Discussion

While much of our analysis above deals with inefficiencies, our results about corporate prediction markets are largely encouraging. First, we find that forecasts from predictions markets outperform other forecasts available to management, including, in the case of Ford, forecasts that are taken extremely seriously. Second, we find that prediction markets get better with age. In both the Google and Ford Sales markets, initial pricing biases disappeared as our sample progressed. This is consistent with the fact that we find more experienced traders trading against pricing biases and earning high returns, and with the fact that traders who appear unskilled stop participating.

It is also consistent with our best and worst-calibrated prediction markets being the markets at Firm X and the Features market at Ford, respectively. The Firm X markets ran for almost 5 years and the average participant made over 200 trades. Our analysis of the Ford Feature markets only covers the first 15 days, when the market appeared to badly misforecast the popularity of features that received only lukewarm support from employees. Unfortunately, given the change to a beauty contest format, we are unable to analyze whether this market too would have improved with age, although the results for the Ford Sales market, which involved an overlapping population of traders, are encouraging.

Regarding the inefficiencies, some results match well with the prior literature, while others are more puzzling. Our finding of an optimistic bias in some markets is consistent with prior work on the role of optimism in organizations. At Google, the optimistic bias is strongest for markets on project completion. Insiders and their friends contribute trade optimistically at Google, potentially for strategic reasons, but also potentially due to overconfidence in one's own and teammates' ability. The fact that the optimistic bias exhibits mood swings is more consistent with a behavioral source. The fact that newly hired employees are the most optimistic is consistent with employees arriving at Google initially miscalibrated and then learning. The fact the optimistic bias diminishes over time is also consistent with initial miscalibration and learning, as it is not obvious that strategic reasons for biases disappeared from 2005 to 2007. Taken together, the evidence suggests that strategic biases, overconfidence, behavior biases, and inexperience (i.e., being a career with systematically erroneous priors) all play a role in the optimistic bias.

The bias in pricing away from naïve priors in Google and Ford's Sales markets is less consistent with prior literature. Most of the extant literature, such as the Ali (1977) and Manski (2006) models, the partition dependence literature, and the work on probability misperceptions (Kahneman and Tversky, 1979), led us to expect a bias in the other direction. We also expected the Inkling market making mechanism to impart a bias towards the prior, at least early in the life of a market, and likewise the potential convexity of reputational incentives should have made low priced securities more attractive, creating a bias in the opposite direction. The fact that the bias away from the prior was strongest at Google (which had the most linear incentives and did not use an automated market maker) was consistent with these expectations, but the overall sign of the bias was not. The pricing bias we did find (at Google and in Ford's Sales markets) is consistent with an overreaction to new information or with participants underappreciating the effort that was put into security design (i.e., insufficient partition dependence). While we still find the direction of the bias puzzling, it did diminish over time, consistent with participants becoming better calibrated. By the end of the sample, there was no evidence of pricing inefficiencies in any of the Google, Ford Sales, and Firm X markets.

Producing efficient forecasts than improved upon the available alternatives was only one of the goal the companies had for their markets. Google's management sought to communicate the importance of its OKRs. The anecdote described above, where a senior manager admitted to having been embarrassed by prediction market trading into a redoubling of efforts, provides at least one example of this working.[20]

We are aware of contrasting anecdotes from other companies. For example, we are aware of four cases at different companies in which internal prediction markets were shut down or limited at the request of senior management after they forecast problems with projects. One of these projects became a high-profile debacle that we believe most readers would be aware of (but which unfortunately we cannot name).

One might regard the closing off of a source of information on a crucial project to be a puzzle, but we can think of several reasons why top management may have wanted to limit the use of prediction markets. The first is that prediction markets in these examples provided an ex ante measure of a key project's quality, where only ex post measures (e.g., market acceptance) would have been available otherwise. Agents, including CEOs, may prefer noisier measures of performance, especially if performance is expected to be disappointing. A second possibility is that top management was already aware of the issues with the project, and so the main contribution of the market would have been to spread this information among rank-and-file employees. Relatedly, the existence of a prediction market may help "cure" employees of their optimism bias, and this may not be in the company's best interests for reasons discussed in the introduction (e.g., in the work of Hoffman, 2011, or Leider and Larkin, 2012). A third possibility is that aggregating information about the imminent failure of a key initiative would have created inside information, creating the need to either limit the trading of employees or to make an early disclosure of the problems to outside shareholders. Google faced exactly this tension in deciding which OKRs could be covered by markets, and ultimately decided against running markets on those OKRs with most relevance to their share price (e.g., ad auction revenue).

In conclusion, we argue that the key question for a firm that is considering implementing prediction market is not whether the markets will produce reasonably well-calibrated forecasts that improve upon existing ones. Our results, from a variety of contexts, suggest that they will, at least after

---

[20] We originally hoped to produce more systematic evidence on this point by randomizing which OKRs were covered by markets, in order to test whether the existence of a market had a causal effect on a project's outcomes. Unfortunately, power calculations revealed that given the number of OKRs at Google for which it was feasible to run markets, the causal effect would have to be implausibly large to be detectable. If this was true at Google, which is among the largest corporate prediction markets run to date, it is likely to be an issue in many other settings.

participants gain sufficient experience. We believe the key question is whether managers will like organizational implications of the information sharing that necessarily accompanies prediction markets. This question is less one of information aggregation in markets and more a question of optimal information sharing within organizations. These questions represent a natural next step for research in this area.

Bibliography

Aggarwal, Rajesh K., and Guojun Wu. "Stock Market Manipulations*." *The Journal of Business* 79.4 (2006): 1915-1953.

Ali, Mukhtar M. "Probability and utility estimates for racetrack bettors." *The Journal of Political Economy* (1977): 803-815.

Allen, Franklin, and Douglas Gale. "Stock-price manipulation." *Review of Financial Studies* 5.3 (1992): 503-529.

Berg, Joyce, et al. "Results from a dozen years of election futures markets research." *Handbook of experimental economics results* 1 (2008): 742-751.

Bergman, Nittai K., and Dirk Jenter. "Employee sentiment and stock option compensation." *Journal of Financial Economics* 84.3 (2007): 667-712.

Chen, Kay-Yut, and Charles R. Plott. "Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem."*California Institute of Technology Social Science Working Paper* 1131 (2002).

Cowgill, Bo, and Eric Zitzewitz. "Mood Swings at Work: Stock Price Movements, Effort and Decision Making." (2008).

Fair, Ray C., and Robert J. Shiller. "The Informational Content of Ex Ante Forecasts." *The Review of Economics and Statistics* (1989): 325-331.

Forsythe, Robert, et al. "Anatomy of an experimental political stock market."*The American Economic Review* (1992): 1142-1161.

Fox, Craig R., and Robert T. Clemen. "Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior."*Management Science* 51.9 (2005): 1417-1432.

Goldstein, Itay, and Alexander Guembel. "Manipulation and the allocational role of prices." *The Review of Economic Studies* 75.1 (2008): 133-164.

Grossman, Sanford J. (1976) “On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information.” *Journal of Finance*, 31(2), 573–585.

Grossman, Sanford J., and Joseph E. Stiglitz. "On the impossibility of informationally efficient markets." *The American Economic Review* 70.3 (1980): 393-408.

Gillen, Benjamin J., Charles R. Plott, and Matthew Shum. "Inside Intel: Sales Forecasting using an Information Aggregation Mechanism." (2013).

Gjerstad, Steven. "Risk Aversion, Beliefs, and Prediction Market Equilibrium." (2005).

Goldstein, Itay, and Alexander Guembel. "Manipulation and the allocational role of prices." *The Review of Economic Studies* 75.1 (2008): 133-164.

Hankins, R., and Alison Lee. "Crowd sourcing and prediction markets." *CHI*. Vol. 11. 2011.

Hansen, Jan, Carsten Schmidt, and Martin Strobel. "Manipulation in political stock markets–preconditions and evidence." *Applied Economics Letters* 11.7 (2004): 459-463.

Hanson, Robin D. "Decision markets." *Entrepreneurial Economics: Bright Ideas from the Dismal Science* (2002): 79-85.

Hanson, Robin. "Logarithmic market scoring rules for modular combinatorial information aggregation." *George Mason University* mimeo (2003).

Hanson, Robin, Ryan Oprea, and David Porter. "Information aggregation and manipulation in an experimental market." *Journal of Economic Behavior & Organization* 60.4 (2006): 449-459.

Hanson, Robin, and Ryan Oprea. "A manipulator can aid prediction market accuracy." *Economica* 76.302 (2009): 304-314.

Jian, Lian, and Rahul Sami. "Aggregation and manipulation in prediction markets: effects of trading mechanism and information distribution."*Management Science* 58.1 (2012): 123-140.

Kahneman, Daniel, and Amos Tversky. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society* (1979): 263-291.

Milgrom, Paul, and Nancy Stokey. "Information, trade and common knowledge." Journal of Economic Theory 26.1 (1982): 17-27. APA

Manski, Charles F. "Interpreting the predictions of prediction markets." E*conomics Letters* 91.3 (2006): 425-429.

Newman, Alexandra Lee. "Manipulation in Political Prediction Markets." *The Journal of Business, Entrepreneurship & the Law* 3.2 (2012): 1.

Oprea, Ryan, et al. "Can manipulators mislead market observers." *Manuscript, University of California Santa Cruz* (2007).

Ortner, Gerhard. "Forecasting markets—An industrial application." *Universität Wien* (1998).

Ottaviani, Marco, and Peter Norman Sørensen. "Outcome manipulation in corporate prediction markets." *Journal of the European Economic Association*5.2-3 (2007a): 554-563.

Ottaviani, Marco, and Peter Norman Sørensen. "Aggregation of Information and Beliefs in Prediction Markets." (2007b).

Petersen, Mitchell A. "Estimating standard errors in finance panel data sets: Comparing approaches." *Review of financial studies* 22.1 (2009): 435-480.

Prendergast, Canice. "A Theory of 'Yes Men'." The American Economic Review (1993): 757-770.

Rhode, Paul W., and Koleman S. Strumpf. "Historical presidential betting markets." *The Journal of Economic Perspectives* 18.2 (2004): 127-141.

Rhode, Paul W., and Koleman S. Strumpf. "Manipulating political stock markets: A field experiment and a century of observational data." *University of Arizona, mimeo* (2006).

Rigobon, Roberto, and Brian Sack. "The effects of war risk on US financial markets." *Journal of banking & finance* 29.7 (2005): 1769-1789.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. "Information (in)efficiency in prediction markets." *Information efficiency in financial and betting markets* (2005): 366.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. "Partisan impacts on the economy: evidence from prediction markets and close elections." *The Quarterly Journal of Economics* 122.2 (2007a): 807-829.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. "Party Influence in Congress and the Economy." *Quarterly Journal of Political Science* 2.3 (2007b): 277-286.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. "Prediction Markets for Economic Forecasting." *Handbook of Economic Forecasting, Vol 2A*. (2013)

Sonnemann, Ulrich, et al. "Psychological biases affect economic market prices." *Unpublished manuscript, University of Münster, Germany* (2011).

Tziralis, George, and Ilias Tatsiopoulos. "Prediction markets: An extended literature review." *The journal of prediction markets* 1.1 (2012): 75-91.

Wolfers, Justin, and Eric Zitzewitz. *Prediction markets*. No. w10504. National Bureau of Economic Research, 2004.

Wolfers, Justin, and Eric Zitzewitz. *Five open questions about prediction markets*. No. w12060. National Bureau of Economic Research, 2006.

Table 1. Summary of corporate prediction markets at Google, Ford, Firm X and selected other companies

Market Topic	Example	Google	Ford	Firm X	Other companies running similar markets
Company performance					
Demand forecasting	Ford F-150 sales next week	X	X		Arcelor Mittal, Best Buy, Chrysler, Eli Lilly, HP, Intel, Nokia
Project completion	Will chat be launched within Gmail by end of quarter?	X			Best Buy, Electronic Arts, Eli Lilly, Microsoft, Nokia, Siemens
Product quality	Google Talk sound quality rating	X			Electronic Arts, Eli Lilly, other pharma
External events	Spandex Price in China	X		X	Eli Lilly, other pharma
Decision markets	If feature X is offered, what will demand be?	X	X		Best Buy, GE, Motorola, pharma, Qualcomm, Rite Solutions, Starwood
Fun	Will the Interns win at the Firm X company picnic?	X		X	Electronic Arts

Incentive Type	Example	Google	Ford	Firm X	Other companies running similar markets
Monetary Prizes	$1 cash awards, credit in company store	X			Best Buy, Microsoft, Misys
Non-Monetary Prizes	T-shirts, plaques	X	X		Microsoft, Misys, other pharma
Reputational Incentives Only	Leaderboard		X	X	Boing, J&J, Microsoft

Market Mechanism	Example	Google	Ford	Firm X	Other companies running similar markets
Decentralized					
Continuous Double Auction		X			Hewlett Packard, Nokia, Siemens
Centralized					
Market Maker (following Hanson (2002) approach)			X	X	Chevron, CNBC, Electronic Arts, GE, General Mills,Lockheed Martin, Microsoft, Missile Defense Agency, Misys, MITRE Corp, Motorola, NASA, Nucor, Overstock.com, PayPal, Proctor and Gamble, Qualcomm, SanDisk, T-Mobile
Other market maker					Boeing, Electronic Arts, Genentech, Hallmark, J&J, Overstock, Sony, WD-40

Approach to Beauty Contest Markets	Example	Google	Ford	Firm X	Other companies running similar markets
Avoided	What will be the price of oil in 2020? Trades resolved according to price of oil in 2020	X		X	Best Buy, Electronic Arts, Google, Microsoft
Included At Least Some	What will be the price of oil in 2020? Trades resolved according to the consensus in the prediction market in July 2012.		X		GE, Motorola, Rite-Solutions

Information about prediction markets run by firms outside of our sample come from public comments by firms and interviews. In some cases, the firm asked not to be identified, or provided only partial information. We omit some examples we are aware of for brevity. It is of course possible that firms have run markets we are unaware of. Note that some companies are listed twice within a section in cases where they changed approaches.

Table 2. Summary statistics on prediction markets in our sample

	Google	Ford	Firm X
Industry	Software/Internet	Automobile	Basic materials
Ownership	Public (Ticker: GOOG)	Public (Ticker: F)	Private
Sample begins	April 2005	May 2010	March 2008
Sample ends	September 2007	December 2010	January 2013
Markets (questions)	270	149	1,345
Securities (answers)	1,116	149	4,278
Trades	70,706	9,258	12,655
Unique traders	1,465	902	57
Market mechanism	IEM-style CDA	LMSR	LMSR
Software	Internally developed	Inkling	Inkling
Style of market			
One continuous outcome (e.g., how many F-150s sold?)		68%	1.3%
One binary outcome (e.g., Project X done by Sep 30?)		32%	59%
Two outcomes (e.g., Yes and No securities)	29%		0.7%
3+ outcomes (e.g., bins)	71%		39%
Topic of market			
Demand forecasting	20%	68%	
Project completion	15%		
Product quality	10%		
External news	19%		96%
Decision	2%	32%	
Fun	33%		4%
Share for which optimism can be signed	58%	100%	71%

Notes: IEM-style CDA = continuous double auction with separate securities for each outcome (Forsythe, et. al., 1992)

LMSR = Logarithmic Market Scoring Rule (Hanson, 2003)

Table 3. Markets vs. Experts

Company	Ford	Google	Firm X	
Prediction market type	One continuous outcome	3-5 bins	3-10 bins	One binary outcome
Expert forecast source	Expert forecaster	Derived from Bins	Derived from Bins	Contract over/under
Market topic	Auto sales	Demand	Macro numbers	Macro numbers
Timing	Simultaneous	Within one day	Within one day	Within one day
Prediction market forecast	0.67	1.12	1.01	1.16
	(0.10)	(0.12)	(0.19)	(0.19)
Expert forecast	0.38	-0.38	-0.11	-0.27
	(0.08)	(0.17)	(0.57)	(0.17)
Observations	78	257	1330	748
Unique markets	6	52	185	296
Time periods	13	10	45	58
MSE (prediction market)/MSE(expert)	0.742	0.727	0.924	0.908
P-value of difference with 1	0.104	0.00004	0.002	0.002

This table presents regressions of the outcome being forecast on forecasts from prediction markets and experts. For Google and Firm X, the expert forecasts are derived from the prediction market security construction as described in the text.

Table 4: Tests for Pricing Biases

Dependent variable: Returns to expiry

Panel A. Google

	(1)	(2)	(3)	(4)	(5)
(Price - Naïve Prior)		-0.232***	-0.259**	-0.226**	-0.211**
		(0.089)	(0.101)	(0.090)	(0.089)
(Price - Naïve Prior)^2			0.104		
			(0.174)		
Optimism				-0.103**	-0.104**
(+1 if best outcome, -1 if worst)				(0.041)	(0.041)
Extreme outcome					0.043*
Abs(Optimism)					(0.023)
Constant	-0.017***	-0.010**	-0.016	-0.006	-0.022***
(Captures short aversion)	(0.004)	(0.004)	(0.012)	(0.004)	(0.008)
Trades	37,910	37,910	37,910	37,910	37,910
Securities	612	612	612	612	612
Markets	157	157	157	157	157
R-squared	0.000	0.025	0.026	0.067	0.074

Panel B. Firm X

	(1)	(2)	(3)	(4)	(5)
(Price - Naïve Prior)		0.026	0.063	0.017	0.003
		(0.050)	(0.050)	(0.050)	(0.053)
(Price - Naïve Prior)^2			-0.243**		
			(0.120)		
Optimism				0.021	0.026
(+1 if best outcome, -1 if worst)				(0.021)	(0.022)
Extreme outcome					-0.052
Abs(Optimism)					(0.041)
Constant	-0.003	-0.003	0.007	-0.010	0.032
(Captures short aversion)	(0.013)	(0.013)	(0.014)	(0.014)	(0.030)
Trades	8,910	8,910	8,910	8,910	8,910
Securities	1,704	1,704	1,704	1,704	1,704
Markets	945	945	945	945	945
R-squared	0.000	0.000	0.002	0.001	0.003

Panel C. Ford

	Sales (1)	Sales (2)	Sales (3)	Features (1)	Features (2)	Features (3)
(Price - Naïve Prior)		-0.222	-0.224		0.711***	0.510***
		(0.153)	(0.158)		(0.124)	(0.122)
(Price - Naïve Prior)^2			-0.123			0.866***
			(0.752)			(0.248)
Constant	-0.013	-0.009	-0.008	-0.106**	-0.240***	-0.269***
(Captures optimism and short aversion)	(0.008)	(0.007)	(0.007)	(0.051)	(0.043)	(0.038)
Trades	3,262	3,262	3,262	5,996	5,996	5,996
Securities	101	101	101	48	48	48
Markets	101	101	101	48	48	48
R-squared	0.000	0.237	0.244	0.000	0.147	0.158

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Each observation is a trade; the dependent variable is the percentage point return to expiry (i.e., expiry value - price). For Google, Firm X and the Ford feature markets, the naïve prior is 1/N, where N is the number of outcomes for the market (N = 2 for binary markets). For Ford Sales markets, the naive prior is the initialization price for the market, which appears to have been based on lagged sales. For Google and Firm X markets, outcomes are ordered based on what would be beneficial for company profits -- the best outcome is scaled +1 and the worst is scaled -1. In Ford's markets, trading optimistically always involves buying, and so an optimistic bias cannot be separately identified from a preferrence for buying rather than selling. In the sample size data, a security refers to a unique security with a specific payoff and a market refers to a group of securities with related payoffs (e.g., a group of securities tracking mutually exclusive outcomes). Standard errors are heteroskedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X)

Table 5: Pricing Biases over the life of markets

Dependent variable: Returns to expiry

Panel A. Google

	Trades 1-2	Trades 3-5	Trades 6-10	Trades 11-25	Trades 26-50	Trades 51-100	Trades 101-200	Trades 201+
(Price - Naïve Prior)	-0.546***	-0.468***	-0.457***	-0.349***	-0.331***	-0.207*	-0.088	-0.037
	(0.067)	(0.086)	(0.102)	(0.079)	(0.082)	(0.106)	(0.154)	(0.187)
Optimism	0.007	-0.004	-0.027	-0.062**	-0.096**	-0.128**	-0.161***	-0.121
(+1 if best outcome, -1 if worst)	(0.034)	(0.033)	(0.035)	(0.030)	(0.038)	(0.052)	(0.059)	(0.090)
Constant	-0.012***	-0.004	-0.006	-0.014**	-0.006	-0.010	0.011	-0.006
(Captures short aversion)	(0.004)	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.012)	(0.015)
Trades	1,094	1,624	2,535	6,257	7,478	8,947	6,921	3,054
Markets	157	154	150	144	112	81	46	17
R-squared	0.081	0.064	0.069	0.059	0.079	0.087	0.132	0.076

Panel B. Firm X

	Trades 1-2	Trades 3-5	Trades 6-10	Trades 11-25	Trades 26+
(Price - Naïve Prior)	0.135	-0.030	0.005	0.055	-0.008
	(0.119)	(0.074)	(0.059)	(0.088)	(0.167)
Optimism	0.025	0.023	0.010	0.014	0.094
(+1 if best outcome, -1 if worst)	(0.017)	(0.020)	(0.031)	(0.042)	(0.102)
Constant	-0.006	-0.003	-0.013	-0.029	-0.038
(Captures short aversion)	(0.009)	(0.013)	(0.020)	(0.035)	(0.117)
Trades	3,113	2,674	1,863	1,129	131
Markets	945	742	429	187	12
R-squared	0.003	0.002	0.000	0.004	0.066

Panel C. Ford Sales

	Trades 1-2	Trades 3-5	Trades 6-10	Trades 11-25	Trades 26-50	Trades 51+
(Price - Naïve Prior)	-0.128*	-0.112	-0.109	-0.114	-0.267	-0.811***
	(0.073)	(0.124)	(0.159)	(0.201)	(0.161)	(0.068)
Constant	0.008	-0.004	-0.013	-0.014	-0.007	-0.004***
(Captures optimism and short aversion)	(0.013)	(0.012)	(0.012)	(0.011)	(0.013)	(0.001)
Trades	202	294	461	1,002	725	558
Markets	101	99	96	86	48	20
R-squared	0.051	0.029	0.025	0.024	0.131	0.710

Panel D. Ford Features

	Trades 1-2	Trades 3-5	Trades 6-10	Trades 11-25	Trades 26-50	Trades 51-100	Trades 101+
(Price - Naïve Prior)	6.263***	3.674***	2.478***	1.213***	0.781***	0.720***	0.811***
	(2.086)	(0.630)	(0.577)	(0.246)	(0.182)	(0.192)	(0.194)
Constant	-0.036	-0.140**	-0.187***	-0.218***	-0.256***	-0.280***	-0.307***
(Captures optimism and short aversion)	(0.066)	(0.056)	(0.031)	(0.033)	(0.042)	(0.056)	(0.089)
Trades	96	144	240	720	1,126	2,024	1,586
Markets	48	48	48	48	48	46	33
R-squared	0.112	0.183	0.252	0.237	0.145	0.165	0.280

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regressions identical to those in Table 4, Column 4 (for Google and Firm X) or Column 2 (Ford) are presented, except that trades in each security are numbered sequentially and the sample is split according to trade number. Standard errors are heteroskedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X).

Table 6: Biases by Subsample

Panel A: Google

	(1)	(2)	(3)	(4)	(5)
	All	Demand Forecasting	Project Completion	Product Quality	External News
Price - Prior	-0.226**	-0.203	-0.247	-0.186	-0.489**
	(0.090)	(0.133)	(0.175)	(0.145)	(0.207)
Optimism	-0.103**	-0.039	-0.239***	-0.085	0.109**
	(0.041)	(0.046)	(0.068)	(0.083)	(0.052)
Constant	-0.006	-0.012*	-0.008	0.006	-0.003
	(0.004)	(0.007)	(0.006)	(0.012)	(0.009)
Trades	37,910	12,387	11590	5897	6,898
Markets	157	51	38	22	42
R-squared	0.067	0.024	0.211	0.207	0.104

Panel B: Firm X

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	All	Politics	Policy	Stocks	Growth	Jobs	Commodities	Exchange Rates	Eurozone	Energy	Inflation
Price - Prior	0.017	-0.058	-0.081	-0.257	-0.013	-0.453	0.699***	0.129	0.452***	0.412***	0.034
	(0.050)	(0.221)	(0.086)	(0.158)	(0.125)	(0.278)	(0.123)	(0.090)	(0.090)	(0.116)	(0.122)
Optimism	0.021	0.163*	0.034	0.001	0.049	0.175***	-0.019	0.012	-0.050	0.020	-0.095**
	(0.021)	(0.090)	(0.120)	(0.086)	(0.032)	(0.044)	(0.043)	(0.066)	(0.057)	(0.051)	(0.041)
Constant	-0.010	-0.026	-0.004	0.028	-0.019	0.043**	0.019	-0.059	0.002	-0.062**	-0.006
	(0.014)	(0.055)	(0.075)	(0.060)	(0.029)	(0.017)	(0.038)	(0.062)	(0.017)	(0.024)	(0.018)
Trades	8,910	449	382	1,309	2,205	657	290	462	166	492	1,429
Markets	945	35	12	53	425	39	35	46	20	41	93
R-squared	0.002	0.119	0.005	0.018	0.003	0.088	0.143	0.005	0.120	0.055	0.032

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regressions identical to those in Table 4, Column 4 are presented for subsets of the Google and Firm X markets. Only markets for which optimism can be signed are included, and thus "Fun" markets are excluded. Standard errors are heteroskedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X).

Table 7: Optimism and Stock Returns
Dependent variable: Returns to expiry

	(1)	(2)	(3)	(4)	(5)	(6)
Optimism*Google log stock return (t+1)	-0.880	-0.237	-0.315	-0.994	-0.622	-0.802
	(0.719)	(0.654)	(0.666)	(0.683)	(0.602)	(0.560)
Optimism*Google log stock return (t)	-1.162	-0.181	-0.236	0.010	0.240	0.298
	(0.797)	(0.452)	(0.431)	(0.614)	(0.616)	(0.488)
Optimism*Google log stock return (t-1)	-2.060***	-1.335**	-1.306**	-2.567***	-2.072***	-1.376**
	(0.752)	(0.570)	(0.565)	(0.766)	(0.654)	(0.625)
Optimism*Google log stock return (t-2)	-0.690	0.044	0.070	-0.092	-0.030	-0.029
	(0.437)	(0.301)	(0.282)	(0.316)	(0.352)	(0.315)
Topics included	All	All	All	Completion	Completion	Completion
Google stock returns (t+1, t, t-1, t-2)	Y	Y	Y	Y	Y	Y
Interactions of Google stock returns (t+1 to t-2) with calendar quarter fixed effects		Y	Y	Y	Y	Y
Interactions of Google stock returns (t+1, t, t-1, t-2) with extremeness and favorites			Y	Y	Y	Y
S&P and Nasdaq returns (t+1, t, t-1, t-2) and interactions with optimism					Y	Y
Day of week fixed effects and interactions with optimism						Y
Observations	37,910	37,910	37,910	11,590	11,590	11,590
R-squared	0.095	0.155	0.159	0.489	0.505	0.522

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The regressions in this table extend the regression in Table 4, Column 5 by adding Google stock returns from surrounding periods and their interaction with the optimism variable. Columns 1-3 include all trades included in Table 4, Column 5 (i.e., all markets for which optimism can be signed), while columns 4-6 include only markets on the timing of project completion (i.e., those included in Table 6, Column 3). Standard errors are heteroskedasticity-robust and allow for clustering within markets and calendar months.

Table 8. Reduction in Biases Over Time

	Google				Ford Sales		Firm X			
	(1)	(2)	(3)	(4)	(1)	(2)	(1)	(2)	(3)	(4)
Constant	-0.011*	-0.001	-0.008	-0.003	-0.036	0.018	-0.056**	-0.057**	-0.018	-0.050
	(0.006)	(0.008)	(0.012)	(0.004)	(0.025)	(0.028)	(0.026)	(0.026)	(0.034)	(0.036)
Constant*Market Start Date (Min 0, Max 1)	-0.007	-0.011	-0.006	-0.010	0.036	-0.031	0.141***	0.143***	0.031	0.077
	(0.013)	(0.014)	(0.005)	(0.012)	(0.036)	(0.044)	(0.040)	(0.039)	(0.052)	(0.053)
(Price - Prior)		-0.390**	-0.295**	-0.349***		-2.158***		0.069	0.113	0.074
		(0.164)	(0.149)	(0.129)		(0.268)		(0.085)	(0.107)	(0.114)
(Price - Prior)*Market Start Date		0.337*	0.162	0.297		2.531***		-0.123	-0.203	-0.149
		(0.195)	(0.308)	(0.292)		(0.317)		(0.150)	(0.205)	(0.212)
Optimism				-0.206***						0.085*
				(0.059)						(0.051)
Optimism*Market Start Date				0.282**						-0.127
				(0.117)						(0.081)
Markets w/o optimism signed included	Yes	Yes	No	No	N/A	N/A	Yes	Yes	No	No
Trades	70,706	70,706	37,910	37,910	3,262	3,262	12,655	12,655	8,910	8,910
R-squared	0.000	0.027	0.026	0.090	0.006	0.322	0.007	0.007	0.001	0.006

Regressions identical to those in Table 4, Columns 1, 2 and 4 are presented, with the variables interacted with a linear time trend, which is scaled to equal 0 at the beginning of the sample and 1 at the end. Columns 3 and 4 for Google and Firm X include only those markets for which optimism can be signed. Standard errors are heteroskedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X).

Table 9. Biases, Experience and cumulative returns

Dependent variable: Security characteristic*(+1 if buy, -1 if sell)

Panel A: Google

	(1)	(2)	(3)	(4)
	Optimism	Price - Prior	Buy	Returns
Cumulative Returns	-1.516***	-0.191**	-0.584***	1.699***
	(0.270)	(0.078)	(0.223)	(0.096)
Experience (Log Trades)	-0.009	-0.031***	-0.119***	0.000
	(0.009)	(0.003)	(0.020)	(0.002)
Observations	75,820	141,412	141,412	141,412
R-squared	0.034	0.053	0.056	0.152

Panel B: Ford Sales

	(1)	(2)	(3)
	Buy/Optimism	Price - Prior	Returns
Cumulative Returns	-0.204	-0.032***	0.025***
	(0.139)	(0.006)	(0.006)
Experience (Log Trades)	-0.117***	-0.008***	0.006***
	(0.023)	(0.001)	(0.002)
Observations	2,810	2,810	2,810
R-squared	0.023	0.040	0.019

Panel C: Firm X

	(1)	(2)	(4)	(5)
	Optimism	Price - Prior	Buy	Returns
Cumulative Returns	-8.135	0.992	8.463	6.732***
	(5.320)	(1.681)	(9.194)	(2.368)
Experience	0.003	0.019***	-0.061*	-0.012***
	(0.011)	(0.005)	(0.035)	(0.004)
Observations	8,696	12,318	12,318	12,318
R-squared	0.001	0.018	0.010	0.005

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

This table presents regressions testing whether traders with more past experience or higher past returns trade in a direction that is correlated which security characteristics or with future returns. In Google's markets, each trade has two participants (a buyer and a seller), and thus each trade appears in the dataset twice. For Ford and Firm X, participants trade with an automated market maker, and so each trade appears in the data once. For each observation, the dependent variable is a security characteristc multipled by the side (+1 if a buy, -1 if a sell). The dependent variable "Buy" is this side variable; "Returns" is returns to expiry multiplied by side. Standard errors are heteroskedasticity-robust and allow for clustering within participants and markets.

Table 10. Linear probability models predicting participation in next period
Dependent variable: = 1 if trader continues to participate

Market	Google	Ford Sales	Firm X
Time period	Quarter	Week	Month
Time period in dep variable	Q+1	W+3	M+2
Ln(Trades in prior period)	0.102***	0.109***	0.006
	(0.014)	(0.028)	(0.005)
Return in prior period	0.063***	0.223**	-0.007
	(0.024)	(0.087)	(0.027)
Dep. Variable mean	0.36	0.32	0.97
Observations	2,266	689	3,363
Unique traders	1,449	294	57
Time periods included	9	12	59
R-squared	0.140	0.068	0.001

This table presents linear probability regressions predicting participation in a future time period based on an individual trader's performance and trading activity in the current time period. The length of the time period is different in each market due to the different circumstances (Ford's Sales market lasted 13 weeks, so we use weeks as the time period), Google's markets involved a new account for each trader in each quarter, so quarter's are the natural frequency at which to study participation decisions, and months were used at Firm X because most markets at Firm X were resolved at a monthly frequency (e.g., markets on monthly economic numbers). Given these choices, the future period was chosen to ensure that most prior-period markets would be settled by the beginning of the period (this was true for every market at Google and Ford, at Firm X returns from the few unsettled markets were not included in the return measure). Standard errors are heteroskedasticity-robust and allow for clustering on participants and time periods.

Table 11. Trader Characteristics, Biases and Returns (Google)

	(1) Optimism	(2) Price - Prior	(4) Buy	(5) Returns
Market Insider	0.126	0.066*	0.404***	-0.013
	(0.099)	(0.034)	(0.153)	(0.046)
Friend of Insider	0.143**	0.034*	-0.194	-0.027
	(0.068)	(0.019)	(0.180)	(0.019)
Coder / Engineer	-0.015	-0.157***	-0.492***	0.088***
	(0.085)	(0.028)	(0.129)	(0.033)
Hire Date (in Years)	0.067**	-0.016	-0.174***	-0.003
	(0.031)	(0.010)	(0.062)	(0.013)
NYC-Based	-0.150	-0.117***	-0.085	0.067*
	(0.115)	(0.032)	(0.159)	(0.037)
Mountain-View Based	-0.177*	-0.037	-0.119	0.038
	(0.094)	(0.026)	(0.104)	(0.025)
Observations	75,820	141,412	141,412	141,412
R-squared	0.009	0.046	0.065	0.005

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This Table presents regressions analogous to those in Table 9, Panel A, except that traded characteristics are included rather than experience variables. A market insider is a participant on the project covered by the market. Friends of insiders are as indicated by either party on a social networking survey. Standard errors are heteroskedasticity-robust and adjust for clustering within participants and markets.

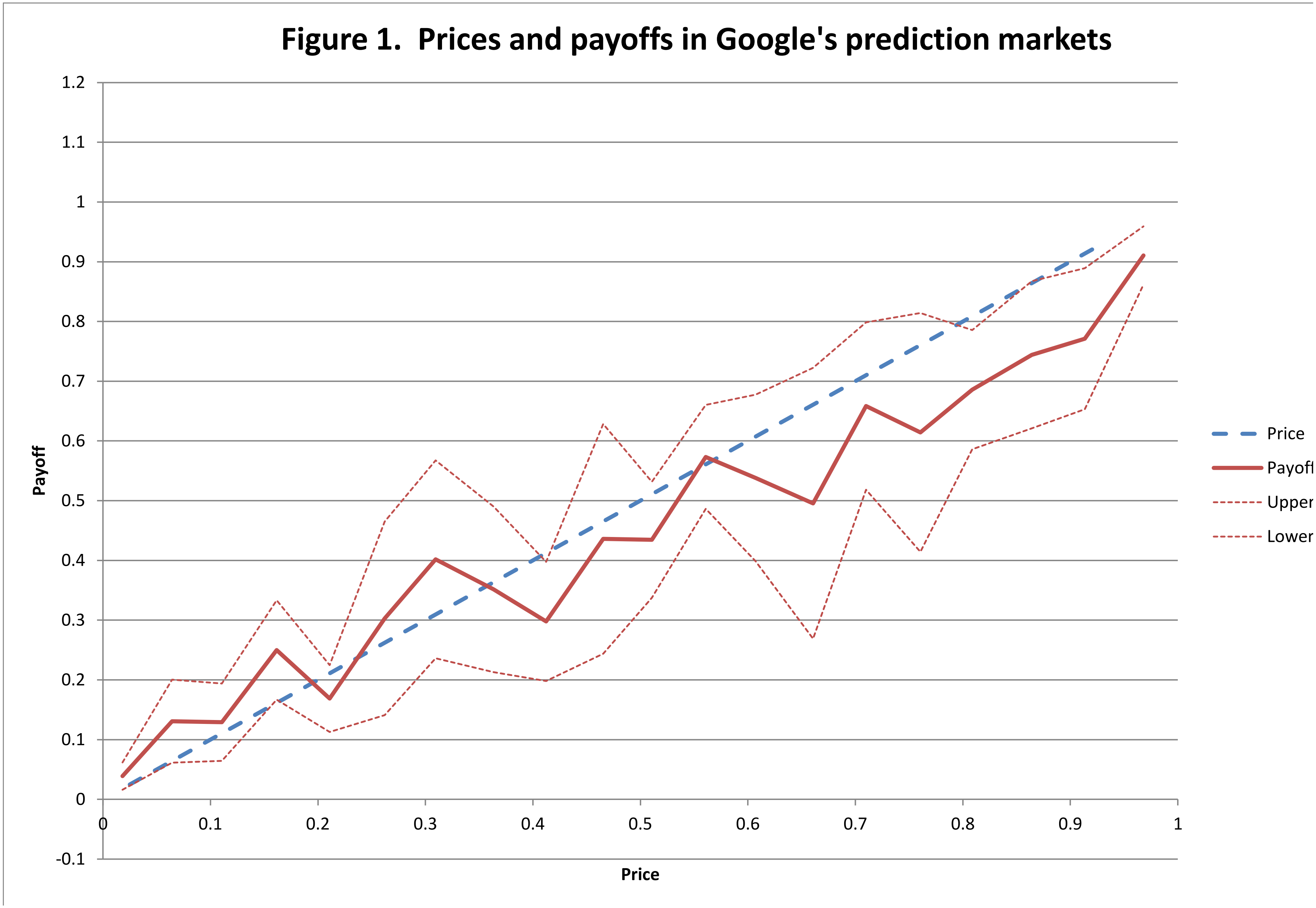


Figure 1. Prices and payoffs in Google's prediction markets
Payoff
Price
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Price
Payoff
Upper
Lower

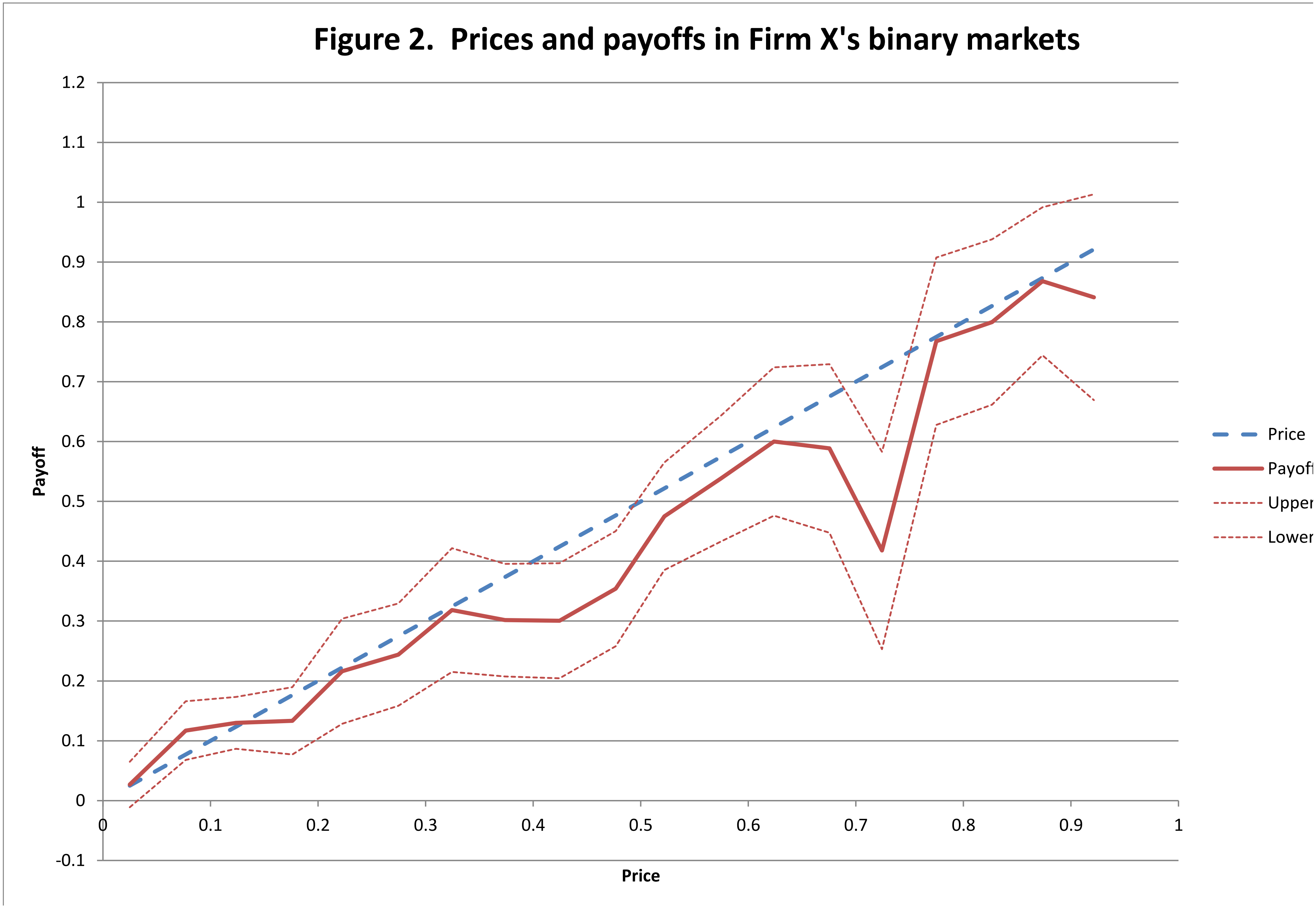
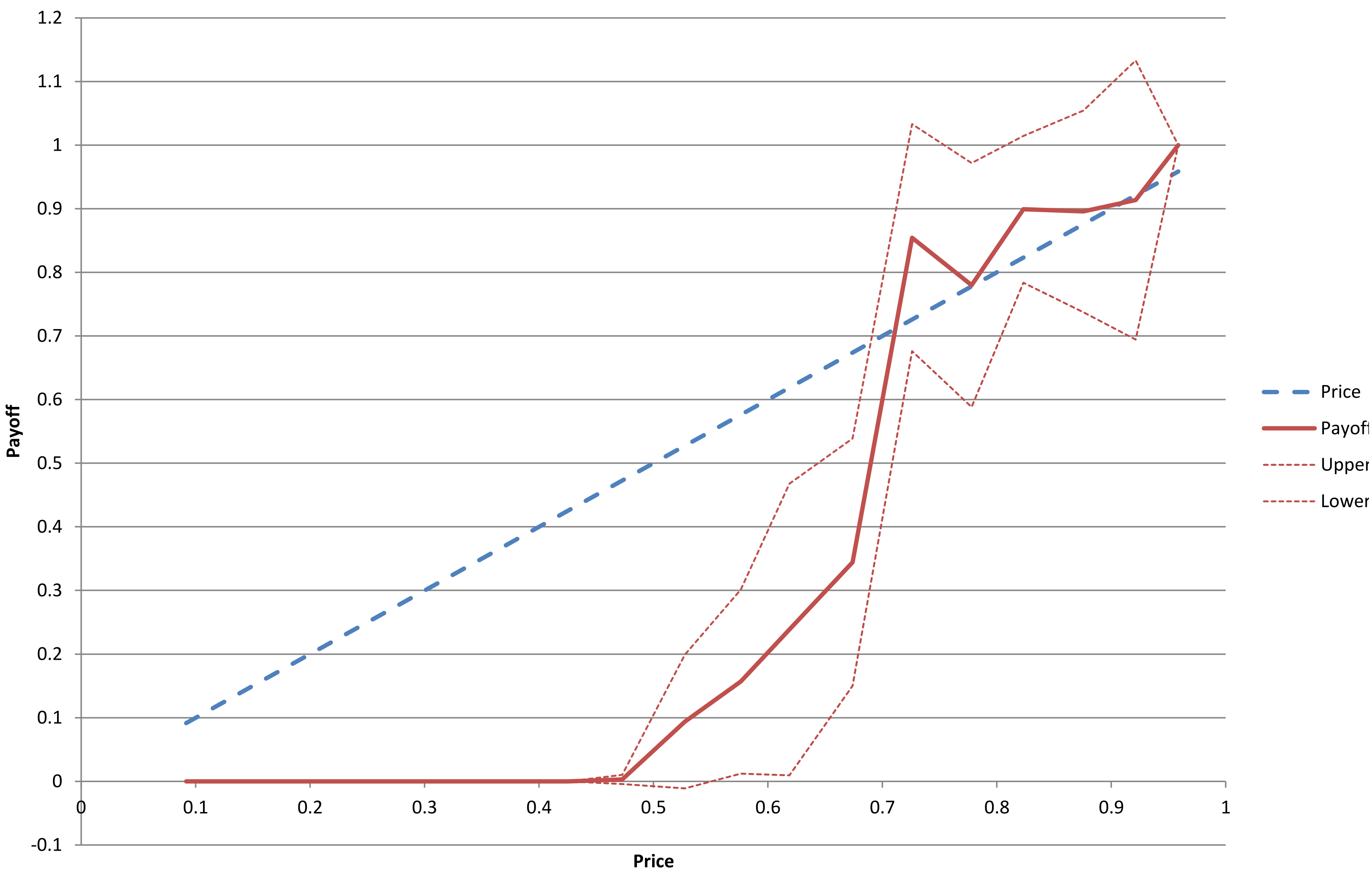


Figure 2. Prices and payoffs in Firm X's binary markets
Payoff
Price
Price
Payoff
Upper
Lower
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 3. Prices and payoffs in Ford's Feature markets



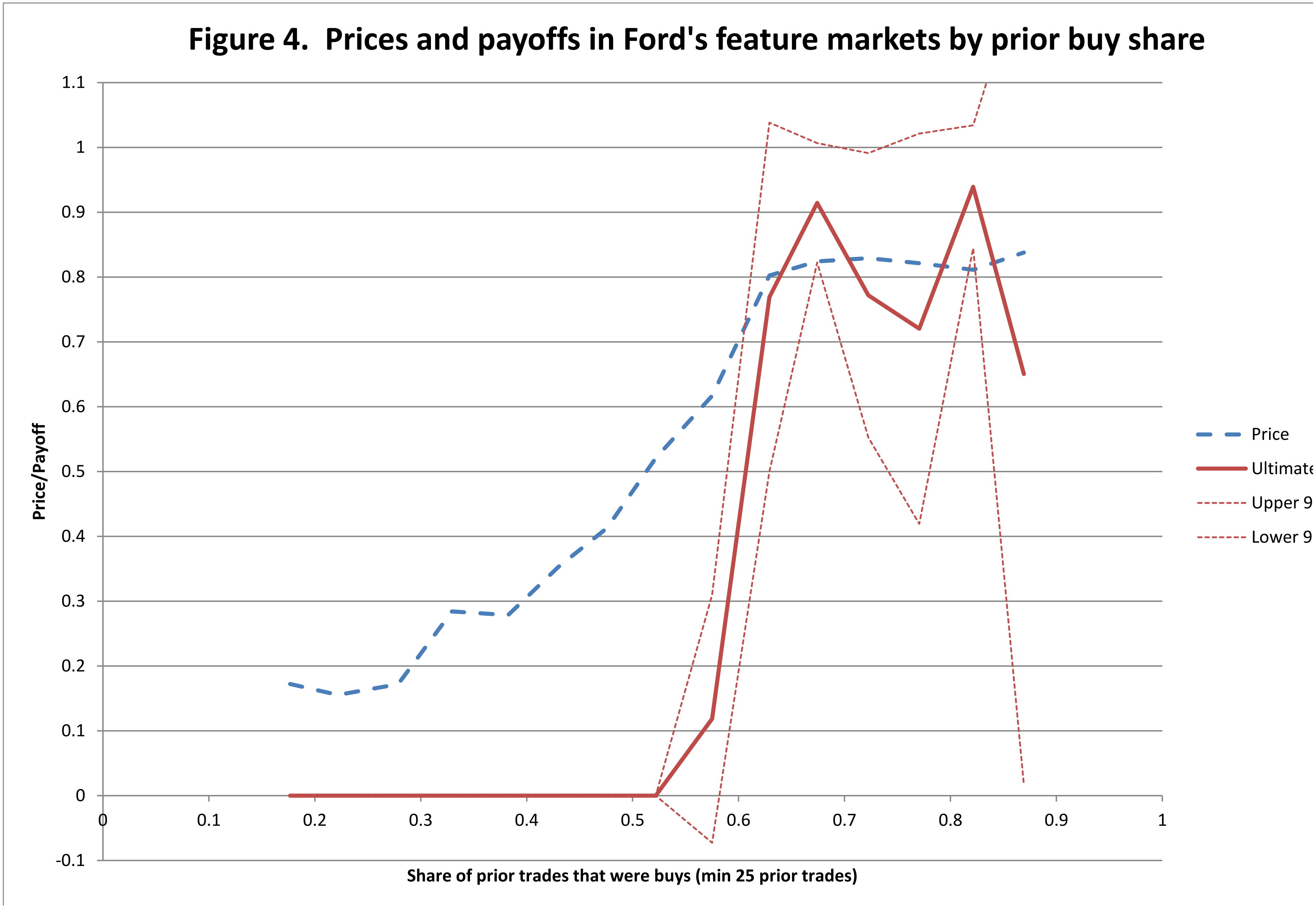


Figure 4. Prices and payoffs in Ford's feature markets by prior buy share
Price/Payoff
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Share of prior trades that were buys (min 25 prior trades)
Price
Ultimate
Upper 9
Lower 9

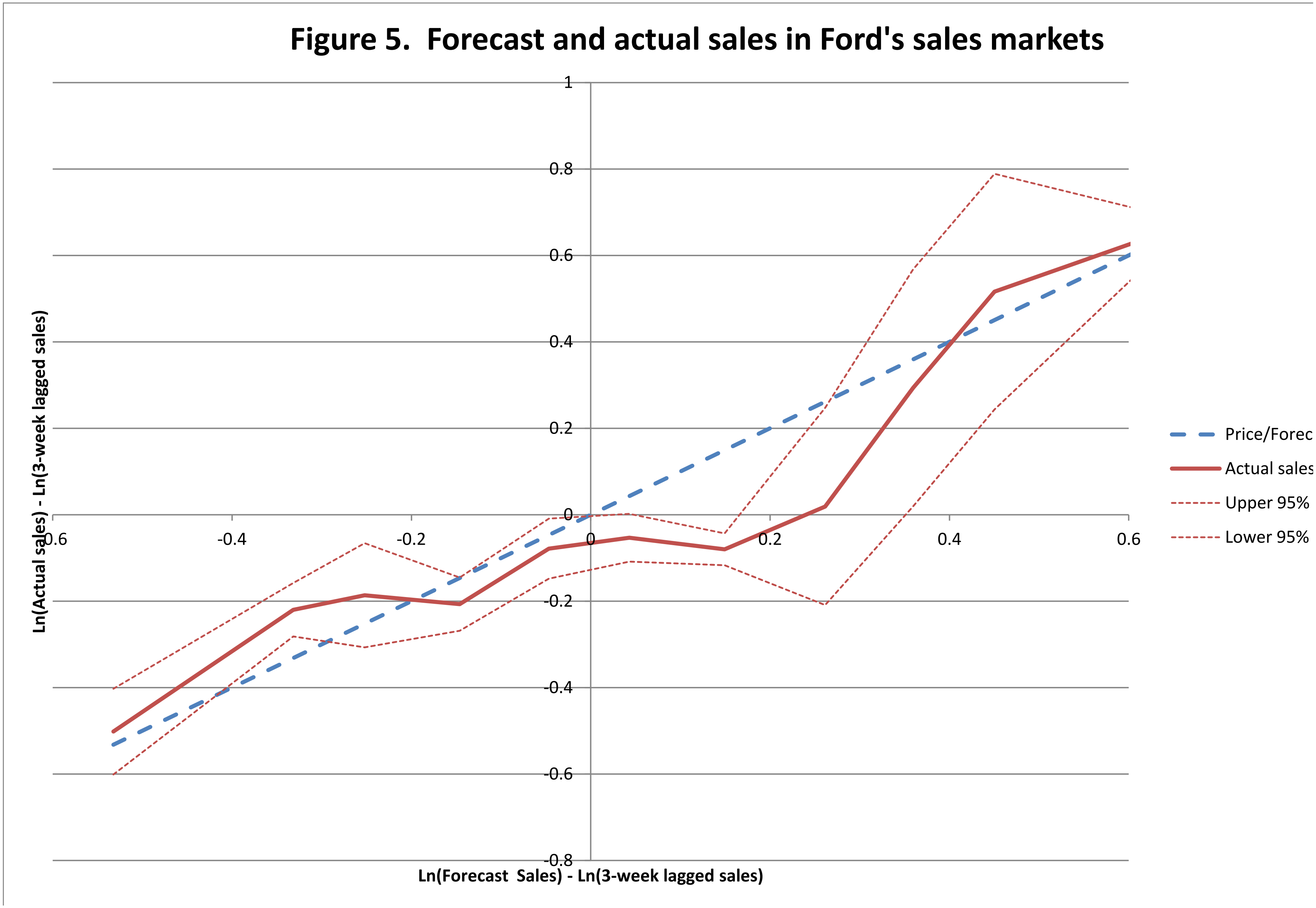


Figure 5. Forecast and actual sales in Ford's sales markets
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Ln(Actual sales) - Ln(3-week lagged sales)
Ln(Forecast Sales) - Ln(3-week lagged sales)
Price/Forec
Actual sales
Upper 95%
Lower 95%