Review of Economic Studies (2015) **82**, 1309–1341 doi:10.1093/restud/rdv014

Advance access publication 2 April 2015

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

BO COWGILL
Haas School of Business, UC Berkeley

and

ERIC ZITZEWITZ
Dartmouth College, CESifo and NBER

First version received July 2010*; final version accepted July* 2014 (*Eds.*)

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 biases or ulterior motives—raising questions about how well these markets will perform. We examine data from prediction markets run by Google, Ford Motor Company, and an anonymous basic materials conglomerate (Firm X). 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. 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.

Key words: Market efficiency, Forecasting, Information sharing, Optimism bias, Overreaction, Underreaction, *Wisdom of Crowds*

JEL Codes: G13, M00

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* (Surowiecki, 2004) was a notable source of enthusiasm.

However, markets in organizations face issues distinct from public prediction markets. If markets are run on topics of strategic importance, there is often a need to limit participation for

*This article 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.

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 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 favourable 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 article examines the efficiency of corporate prediction markets by studying markets at three major companies: Google, Ford Motor Company, and Firm X.[1] 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 ($p = 0.104$). 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 three 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. Both Google and Ford's markets exhibit a bias away from a naive prior ($1/N$, where N is the number of bins, for Google and prior sales for Ford). However, we find that these inefficiencies 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. Secondly, 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, they learn about their ability over time, and 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% of trades, that

1. Firm X is a large, privately held, profitable basic materials and energy conglomerate headquartered in the Midwestern U.S., but with global operations.

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 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 (2003) 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

Notes: 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.

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. Hence, 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, 2015). 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 article 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) analyse 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 *et al.* (2005, 2013) 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 *et al.*, 2007a,b; Wolfers and Zitzewitz, 2009).

While most of the smaller literature on corporate prediction markets is empirical, Ottaviani and Sorenson (2007) 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 *et al.* (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 about 60% as large as the sales forecasting 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 analyse 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. Secondly, the microdata available on Google participants allow us to identify the characteristics of employees who trade with and against inefficiencies. Thirdly, the markets we analyse 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

2. The Intel sales forecasting markets cover 46 product*period combinations, whereas 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 article (Cowgill), but several years before his beginning his career as an economist.

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 horse racing pari-mutuel markets, Manski (2006) shows that two common features of prediction markets—budget constraints and the skewed pay-off 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 (2006a), and Ottaviani and Sorensen (2015) 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, an uncorrected-for bias in information (*e.g.* the "inside view" of Kahneman and Lovallo, 1993) 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 *et al.* (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 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 behavioural, 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, 2003; Compte and Postlewaite, 2004) or lead to risk-taking that generates positive externalities (Bernardo and Welch, 2001; Goel and Thakor, 2008).

Other work has discussed how employee optimism and equity compensation interact. Optimistic employees may overvalue equity compensation, and thus be cheaper to compensate. As Bergman and Jenter (2007) point out, however, the simplest version of this explanation of equity compensation ignores the fact that employees of public companies can buy equity with their cash compensation. Shorting employer equity is difficult for most employees, hence when equity is included in compensation, in practice it likely provides a lower bound on employees' stock exposure. Oyer and Schaefer (2005) argue that firms may use a mixture of equity and cash compensation because it causes employees who are optimistic about firm prospects to self-select into employment, which could be beneficial if optimistic employees work harder, or if they take risks that are beneficial to their employers.

Empirical work finds that employee optimism or overconfidence is correlated with risk-taking, but suggests that the benefits of optimism-induced risk taking may be mixed. Hirshleifer *et al*. (2012) find that firms with overconfident CEOs invest more in research and development and attain more and more highly cited patents. Larkin and Leider (2012) find that overconfident employees select more convex incentives contracts, and Hoffman and Burks (2013) finds that overconfident truckers select more training. In both cases, employee overconfidence lowers costs for firms. At the same time, Malmendier and Tate (2008) find that overconfident CEOs undertake mergers that are associated with lower stock performance for their employers.

4. See Wolfers and Zitzewitz (2004 and 2006b), Rhode and Strumpf (2004 and 2006), Hansen *et al.* (2004), and Newman (2012) for discussions.

Corporate prediction markets provide tools for both measuring and potentially correcting employee optimism. The optimistic bias in Google's markets, and the fact that appears to arise from new employees who become better calibrated with experience, is interesting in light of the aforementioned work. Firm X told us that a primary motivation for running markets was a desire to help senior managers become better calibrated forecasters. It is possible that in their context of economic forecasting and strategic planning, correct calibration is paramount, whereas in other contexts correcting employee optimism may or may not be in an employer's interests.

The remainder of the article 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 article (Cowgill) before beginning 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: (i) aggregating information for management about the success of an important project and (ii) 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 behaviour. 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% 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.[5]

Ford Motor Company is a global automotive manufacturer based in Dearborn, Michigan, with operations and distribution on six continents and a financial services arm called Ford Motor Credit Company. 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 traditional market research, such as focus groups or surveys). Ford limited participation to employees with relevant expertise (in the Marketing and Product Development Divisions).

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 *et al.*, 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 pay-off 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). In a decision market, securities pay-off based on an outcome variable, assuming the decision is undertaken. If the decision tracked by a security is not undertaken, then trades are cancelled. As a result, securities prices should reflect the expected value of the outcome variable, conditional on the decision being undertaken. Rather than defining the outcome as a feature's long-term success in the marketplace, Ford chose feedback from market research as a more immediate outcome measure. Its markets

5. Further detail on Google's prediction markets is available in the original version of this article (Cowgill *et al.*, 2009) and in a Harvard Business School teaching case (Coles *et al.*, 2007).

asked whether a series of potential car features (*e.g*. an in-car vacuum) would reach a threshold level of interest in market research, if that research were conducted.

Traditional market research is expensive to run, sample sizes are necessarily small, and hence sampling errors can be meaningful. In contrast, opinions of employees may be cheaper to obtain, but employees are potentially biased, which is the reason non-employees are consulted in the first place. By asking employees to predict the results of the traditional market research, Ford sought to increase sample sizes while mitigating any biases in employee opinion. 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] Ford also found the qualitative comments market participants made via the prediction market software to be of independent value. Ford cited employee education and engagement as additional benefits of running prediction markets.[7]

Unfortunately for research purposes, shortly after launching the features markets, Ford decided that results of its market research were too sensitive to be shared with its market participants, given the potential for imitation by competitors. As a result, it began settling markets based on the final trade price, rather than the market research outcomes. This turned the markets into "beauty contest" markets, in which security pay-offs depend only on future market prices (Keynes, 1936). While an analysis of the predictive power of these markets would have been interesting, unfortunately this decision also meant the relevant market research outcomes were not recorded in our data, and subsequent attempts to obtain them were unsuccessful. As a result, we have reluctantly omitted them from the analysis.[8]

Firm X is a large, privately held, and profitable diversified basic materials and energy conglomerate headquartered in the Midwestern U.S., but with global operations. 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 macroeconomy and/or to commodity prices, both of which were quite volatile during our sample period (March 2008–January 2013). 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 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 its 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).

6. See http://www.hpcwire.com/2011/02/22/ford_motor_company_turns_to_cloud-based_prediction_market_software/ (last accessed 23 April 2015).

7. Montgomery *et al*. (2013) discusses these additional benefits in more detail.

8. In an earlier version of this article, we analysed a single round of features markets that were run before this change was made. Those markets were poorly calibrated. Markets trading at high prices were roughly efficient, but those trading at low and intermediate prices displayed a very large optimism bias. Features with securities that traded below their initial price never achieved the threshold level of customer interest, and therefore were always expired at zero, and yet the market appeared to not anticipate this. Subsequent discussions with Ford revealed that these markets included features that were not shown to customers, and that these markets may have been unwound rather than expired at zero. Given the uncertainty about returns in these markets, we decided to omit an analysis of these markets from the revised paper, but include a graph documenting the poor calibration of the Features markets in the online appendix.

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.[9]

Just under 60% 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, whereas almost all of the other 80% specified a single "over/under" threshold. A very small number of markets (18 out of 1345) used the linear pay-offs used by Ford's sales markets. For the remaining 40% 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, whereas Ford focused on sales forecasting and decision markets, and Firm X focused on external events. A few other firms have run many types of prediction markets (*e.g.* Eli Lilly, Best Buy), whereas others have run more focused experiments with one particular type of market.[10]

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 pay-offs for the continuous outcomes in its sales markets and single binary securities for the discrete outcomes in its features markets. 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; Sonnemann *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 (*e.g.* Forsythe *et al.*, 1992), in which the range of possible future outcomes is

9. 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.

10. 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.

TABLE 2
Summary statistics

	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	101	1345
Securities (answers)	1116	17	4278
Trades	70,706	3262	12,655
Unique traders	1465	294	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?)		100	1.3
One binary outcome (*e.g.* Project X done by September 30?)			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	100	
Project completion	15		
Product quality	10		
External news	19		96
Decision	2		
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).

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, such as whether a deadline or quality target will be met, there are generally two outcomes, and no reason to expect the *ex ante* 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. In markets with more than two outcomes, this approach does make shorting a security less convenient than taking a long position, since one must either first exchange currency for a complete set of securities and then sell the security, or else buy the securities linked to all other outcomes. On the contrary, any inconvenience cost of shorting should affect all securities in a market at least approximately equally, and biases to prices should be limited by the fact that other participants can simultaneously sell all outcomes if their bid prices sum to more than one. Google did not have an automated market maker, but traders were observed placing such arbitrage trades (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. The presence of the automated market maker also makes shorting or taking long positions equally convenient.

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 price, 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 too 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 the Inkling mechanism could potentially reinforce potential biases towards 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 analyse.[11] 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 pay-offs 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.

11. We have unfortunately been unable to obtain a precise percentage.

2. RESULTS

This section presents statistical tests in four subsections. The first subsection provides simple tests of the calibration of the three firms' markets. We test whether securities are priced at the expectation of their pay-offs, conditional on price alone. The second Section 2 examines whether forecasts from prediction markets improve on contemporaneous expert forecasts. The Section 3 expands our analysis of price efficiency to include tests for an optimism bias and for how pricing biases evolve over time. The final subsection examines how trader skill and experience are related to trading profits 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. *Calibration*

In this subsection, we test whether the markets at Google, Ford, and Firm X make efficient forecasts, in the sense that they do not make forecasting errors that are predictable at the time of the forecast. This is equivalent to asking whether the markets yield 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 $E(Y|H_t)$, where H_t 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, $P_t = E(Y|H_t)$, and expected future returns are zero, $E(Y-P_t|H_t)=0$.

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 this subsection, we begin by asking whether future prediction market returns are correlated with the current price level or the difference between the current price and a naive prior (either the market maker's initial price or $1/N$, where N is the number of mutually exclusive outcomes).

Figures 1 and 2 graph the future value of securities, conditional on current price for binary securities at Google and Firm X, respectively.[12] 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 pay-offs for each bin are graphed on the x- and y-axes, respectively. A 95% confidence interval for the average pay-off is also graphed, along with a 45°line for comparison. The standard errors used to construct the confidence interval are heteroscedasticity robust and allow for clustering within market.[13] Observations are weighted by time-to-next trade, which weights trades according to the amount of time that they persist as

12. All securities in the Google markets are binary and none of the contracts in the Ford sales markets are (they had pay-offs linear in vehicle sales). Almost all Firm X markets are binary—the exceptions were a small number of markets with linear pay-offs in commodity prices. These markets accounted for just under 1% of markets and trades, and they are excluded from Figure 2.

13. 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. For Google and Firm X, we also cluster on calendar month as a second dimension in the regression tables presented below, using the code provided by Petersen (2009). This yields standard errors that are very similar to those that cluster only on market. The Ford prediction markets were short lived enough that we do not have a sufficient number of calendar months for clustering to be valid, hence we instead use one-dimensional clustering on markets (which, in the Sales markets, is also equivalent to clustering on time periods). With only six models in the Ford markets, clustering on model as well would not yield asymptotically valid standard errors.

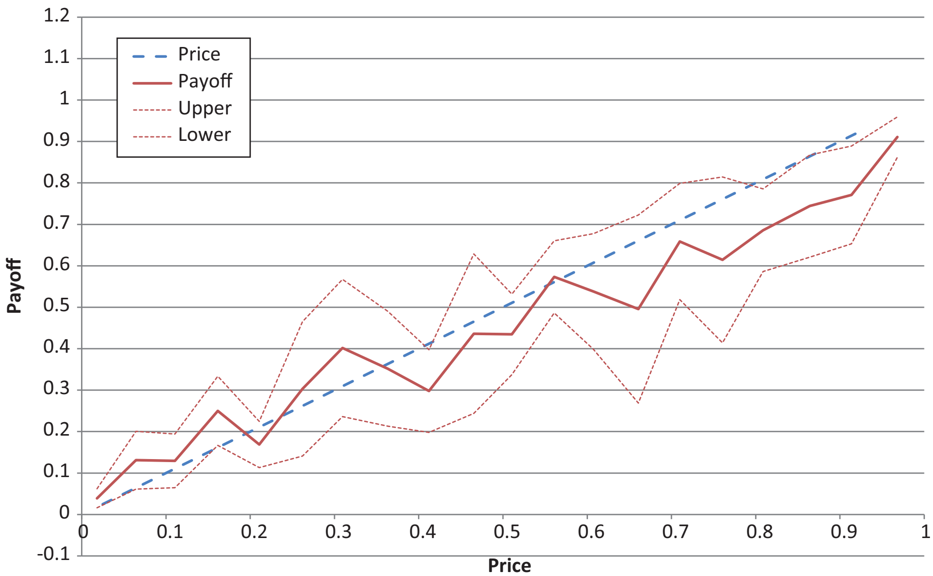

FIGURE 1

Prices and pay-offs in Google's prediction markets.

Trades in Google's prediction markets ($N = 70{,}706$) are sorted into 20 bins based on their price (0–0.05, 0.05–0.01, etc.). The graph plots the average price and ultimate pay-off for each bin. The 95% confidence intervals are reported for pay-offs, based on standard errors that allow for clustering in pay-offs for securities with related outcomes.

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.[14]

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. For Google, the price level below which we observe overpricing differs in two and five-outcome markets. Figure 3A and B plot percentage point returns to expiry (*i.e.* the difference between pay-off and price) against price for Google's two- and five-outcome markets, respectively.[15] In both sets of markets, securities are underpriced when priced below $1/N$ and overpriced when priced above this level, implying a bias in prices away from $1/N$.

Figure 4 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 pay-offs using a model's past sales. To ensure that we do not condition our analysis on information that market participants would not have observed, we use three-week lagged sales. The x-axis plots the log difference between the sales forecast by a trade and lagged sales, and the

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

15. Following other work on binary prediction markets (*e.g.* Tetlock, 2008), we use percentage point returns to expiry (*i.e.* pay-off - price) rather than scaling returns by their price [*i.e.* (pay-off - price)/price]. We do so for two reasons: (i) we are primarily interested in returns as a measure of forecasting performance, rather than financial profit opportunities, and therefore there is no reason to be more interested in a given sized percentage point profit opportunity when the price is low; (ii) scaling by price causes the returns of very low priced securities to dominate the results, and makes the outcome variable more heteroscedastic.

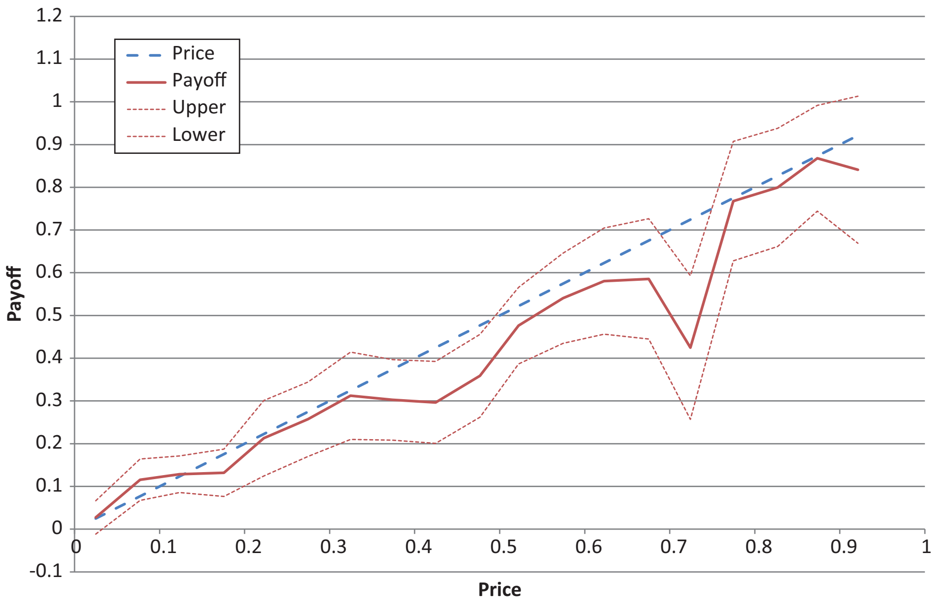

FIGURE 2

Prices and pay-offs in Firm X's binary markets.

Trades in Firm X's binary prediction markets ($N = 9237$) are sorted into 20 bins based on their price (0–0.05, 0.05–0.1, etc.). The graph plots the average price and ultimate pay-off for each bin. The 95% confidence intervals are reported for pay-offs, based on standard errors that allow for clustering in pay-offs for securities with related outcomes.

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 3 presents regressions that test the calibration of the three firms' markets. For each market, we begin with regressions of pay-off on price, where the unit of observation is a trade. If prices are efficient forecasts, then $\mathrm{E}(Y_\mathrm{t}\,|P_t)=P_t$, and a regression of Y_t on P_t should yield a slope of one and a constant of zero. The second regression reported for each market is a regression of percentage point returns to expiry $(Y_t - P_t)$ on P_t. In these regressions, efficient forecasting would be consistent with a slope of zero and a constant of zero. For obvious reasons, the slope in the first regression is simply one plus the slope in the second regression.

The results imply that we cannot reject the null hypothesis of efficient forecasting for the Firm X markets. For the Google markets, we can reject this null, but we still conclude that prices are informative as they are strongly positively correlated with outcomes. For Google, the relationship is slightly less than one-for-one, which implies that high-priced contracts are overpriced and low-priced contracts are underpriced, consistent with Figure 1. For Ford, we cannot reject the hypothesis that the price–outcome relationship is one-for-one, but there is evidence of a negative constant, consistent with a small optimism bias.

Table 3 also reports regressions for Google and Firm X that test whether returns (*i.e.* forecast errors) are better predicted by price or by the difference between price and $1/N$. Whereas the Firm X markets exhibit no predictability with respect to either variable, returns in the Google markets are better predicted by (price $-$ $1/N$) than by price, consistent with Figure 3A and B. We report separate regressions for two- and five-outcome markets, which collectively account for 92% of trades in Google's markets and 65% in Firm X's markets. These regressions suggest

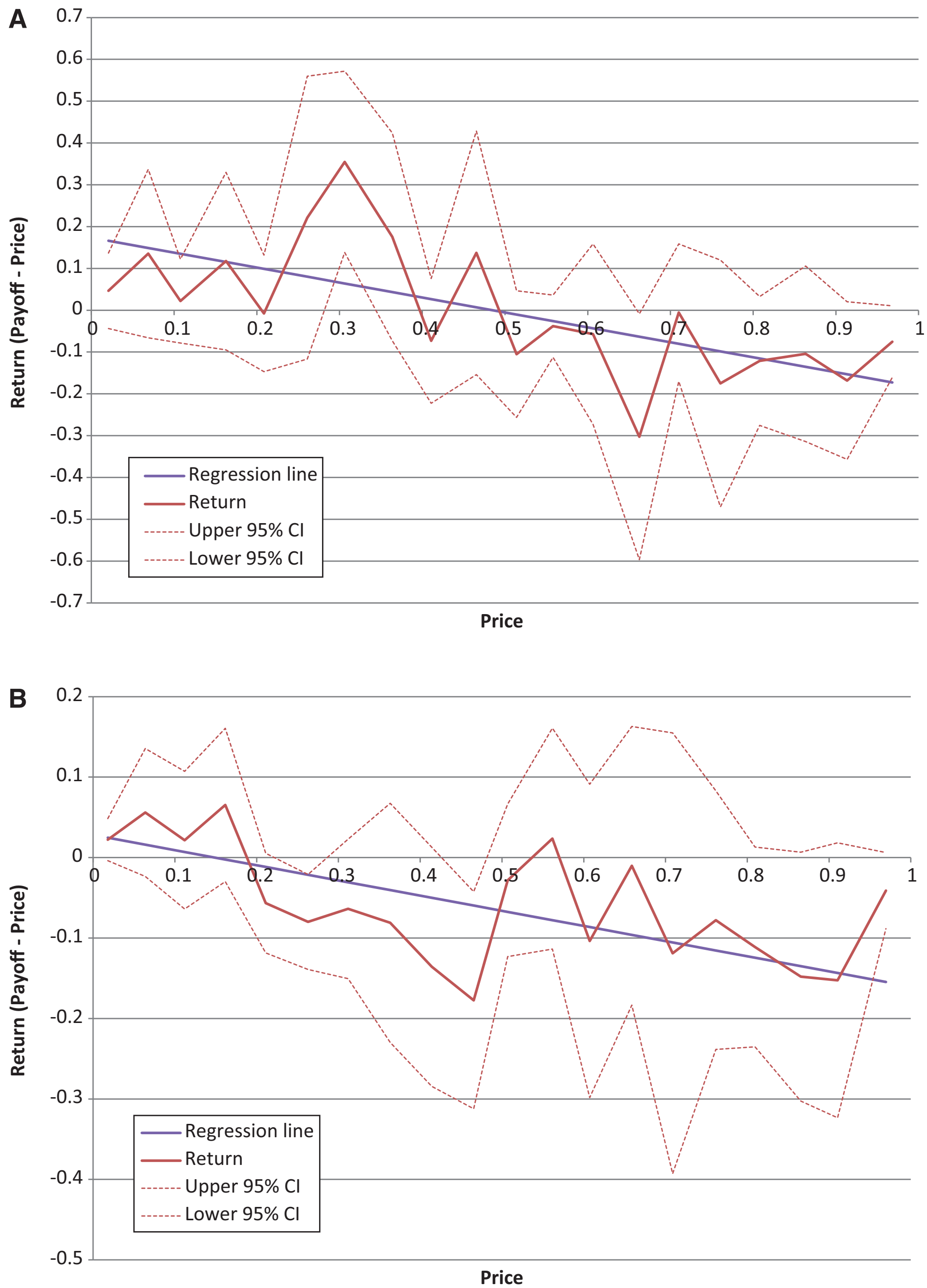

FIGURE 3

Prices and returns in Google's (A) two-outcome and (B) five-outcome markets.

(A) Trades in Google's two-outcome prediction markets ($N = 22{,}452$) are sorted into 20 bins based on their price (0–0.05, 0.05–0.1, etc.). The graph plots the average price and ultimate pay-off for each bin. The 95% confidence intervals are reported for pay-offs, based on standard errors that allow for clustering in pay-offs for securities with related outcomes. (B) Trades in Google's five-outcome prediction markets ($N = 42{,}416$) are sorted into 20 bins based on their price (0–0.05, 0.05–0.1, etc.). The graph plots the average price and ultimate pay-off for each bin. The 95% confidence intervals are reported for pay-offs, based on standard errors that allow for clustering in pay-offs for securities with related outcomes.

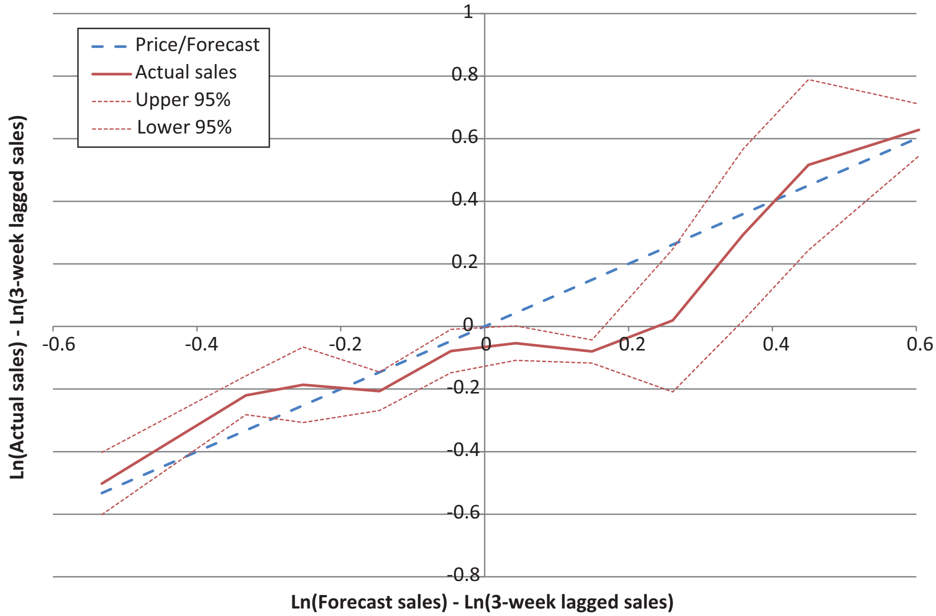

FIGURE 4

Forecast and actual sales in Ford's sales markets.

Trades in Ford's sales prediction markets ($N = 3262$) are sorted into bins based on the log difference between the sales predicted by their price and three-week priors for the given model. The graph plots the average price and ultimate pay-off for each bin. The 95% confidence intervals are reported for pay-offs, based on standard errors allow for clustering in pay-offs for securities for the same time period.

predictability in both types of Google's markets, again consistent with the figures, but neither subset of Firm X's markets. For Ford, we substitute the most recent sales figure reported prior the market commencing as our naive prior, and likewise test whether returns a better priced by (price - prior) sales than by price. We do not find statistically significant evidence that either variable predicts returns for Ford.

Taken together, the results from the figures and Table 3 suggest that all markets have prices that are positively correlated with outcomes, and the Firm X, Ford sales, and Google markets are reasonably well calibrated. While the Firm X and Ford sales markets exhibit no evidence of return predictability, the Google markets display a bias in pricing away from a naive prior of $1/N$. This bias is the opposite of the long shot 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 under reacting to the information used in designing the boundaries or overreacting to other information, such as new information or their own prior beliefs (as in Ottaviani and Sorensen, 2015).

2.2. *Markets versus experts*

Given that firms run prediction markets at least partly to obtain predictions, a natural next 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 a team of expert forecasters. Ford forecasts weekly auto sales

TABLE 3
Calibration tests

Panel A: Google

	(1)	(2)	(3)	(4)	(5)	(6)
Markets included	All	All	All	All	Two-outcome	Five-outcome
Dependent variable	Pay-off	Pay-off − Price	Pay-off − Price	Pay-off − Price	Pay-off − Price	Pay-off − Price
Price	0.812***	−0.188**	0.006			
	(0.083)	(0.083)	(0.024)			
(Price – $1/N$)			−0.238**	−0.232**	−0.357*	−0.189**
			(0.094)	(0.102)	(0.217)	(0.075)
Constant	0.050	0.050	−0.007	−0.006	−0.005	−0.010*
	(0.031)	(0.031)	(0.011)	(0.005)	(0.005)	(0.005)
Trades	70,706	70,706	70,706	70,706	22,452	42,416
Securities	1032	1032	1032	1032	157	767
Markets	270	270	270	270	79	155
Calendar months	30	30	30	30	30	30
R^2	0.255	0.018	0.023	0.023	0.043	0.017

Panel B: Firm X

	(1)	(2)	(3)	(4)	(5)	(6)
Markets included	All	All	All	All	Two-outcome	Five-outcome
Dependent variable	Pay-off	Pay-off – Price	Pay-off – Price	Pay-off – Price	Pay-off – Price	Pay-off – Price
Price	0.969***	−0.031	−0.069			
	(0.069)	(0.069)	(0.117)			
(Price – $1/N$)			0.080	0.021	0.062	−0.018
			(0.113)	(0.046)	(0.054)	(0.097)
Constant	0.028	0.028	0.040	0.017	−0.029	0.032***
	(0.025)	(0.025)	(0.039)	(0.011)	(0.019)	(0.009)
Trades	12,655	12,655	12,655	12,655	5702	2570
Securities	2801	2801	2801	2801	825	782
Markets	1345	1345	1345	1345	818	195
Calendar months	59	59	59	59	59	49
R^2	0.286	0.0004	0.0013	0.0001	0.0010	0.0001

Panel C: Ford Sales

	(1)	(2)	(3)	(4)
Dependent variable	Pay-off	Pay-off – Price	Pay-off – Price	Pay-off – Price
Price	1.046***	0.046	0.057	
	(0.032)	(0.032)	(0.035)	
(Price – Prior sales)			−0.238	−0.222
			(0.144)	(0.153)
Constant	−0.026***	−0.026***	−0.025***	−0.009
	(0.004)	(0.004)	(0.006)	(0.007)
Trades	3262	3262	3262	3262
Securities	101	101	101	101
Markets	17	17	17	17
R^2	0.922	0.022	0.126	0.092

Notes: Standard errors in parentheses. Each column in each panel represents a regression; each observation in these regressions is a trade. The dependent variable is either the ultimate pay-off of a security or the difference between this pay-off and the trade price as indicated. The independent variable is either the trade price or the difference between the trade price and a proxy for a naive prior. For the Google and Firm X markets, in which each market consists of securities linked to N mutually exclusive outcomes, we use $1/N$ as the naive prior probability for each outcome. For the Ford markets, in which each security has a pay-off that is a linear function of the sales of a particular group of models in a given week or month, the naive prior is that most recent actual sales figure reported as of the beginning of the market in question. Standard errors are heteroscedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X). $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$.

for different models, and for the six models covered by prediction markets, we can compare the experts' forecast with the prediction market forecast from immediately before the forecast was issued.[16]

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, to avoid minimize pricing 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.[17]

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% *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 analyse only over/under markets where this approach was *not* used to focus on instances where an over/under value was actively selected.

We compare forecasts that are as close to contemporaneous as possible. For Ford, prediction markets were begun several days before the expert forecast was made, hence we were able to compare the expert forecast with the prediction market forecast immediately before the expert's forecast. For Google and Firm X, the expert forecast was used to design the securities, and therefore it was necessarily made a few days before the prediction market was opened. To limit the timing difference between the expert and prediction markets forecasts, we use prediction markets forecasts from only the first day that a market was open. The prediction market traders may have had access to a few days of information that was not yet available to the expert at Google and Firm X, but this should not have been the case for Ford.

Table 4 presents the results of these comparisons. In each column, we report the results of horse race regressions (Fair and Shiller, 1989) of the security pay-offs on the prediction market and expert forecasts. We also report the ratio of the prediction market and expert meansquared 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 horse race 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.

16. 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.

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

TABLE 4
Markets versus 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 of prediction market forecast	Just before expert	First day of PM	First day of PM	First day of PM
Prediction market forecast	0.67	0.82	1.01	1.16
	(0.10)	(0.14)	(0.19)	(0.19)
Expert forecast	0.38	0.09	−0.11	−0.27
	(0.08)	(0.58)	(0.57)	(0.17)
Observations	78	197	1330	748
Unique markets	6	191	185	296
Time periods	13	30	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

Notes: Standard errors in parentheses. This table presents horse race regressions of the outcome being forecast on forecasts from prediction markets and experts. As described in the text, the prediction market and expert forecasts are as contemporaneous as possible. Standard errors are heteroscedasticity robust and allow for clustering on market and, for Google and Firm X calendar month. In the bottom of the panel, the ratio of the mean-squared errors of the two forecasts is reported. For Ford, the expert forecast is a formal expert forecast, whereas for Google and Firm X the expert forecasts are derived from the prediction market security construction as described in the text.

Nevertheless, it is interesting to note that the mean-squared error improvement achieved by the prediction market at Ford is among the largest.[18]

2.3. *Prediction market pricing biases*

This subsection expands our earlier analysis of prediction market efficiency. In particular, we test whether forecasting errors can be predicted by a broader set of variables than price alone.

In Tables 5 and 6, we test for an optimism bias by adding a variable that captures whether a security is linked to an outcome that we judge would be good for the company. In the Ford sales markets, all securities are structured so that buying involves an expression of optimism (*i.e.* predicting high sales), and thus it is impossible to distinguish between optimism and a preference for taking long rather than short positions. In Google and Firm X's markets, however, securities were available that were linked to both positive and negative outcomes, and hence we can separate these two effects. In these markets, we code the most optimistic outcome as $+1$, the least optimism as -1, and place intermediate outcomes at uniform intervals along this scale (*e.g.* in 5-outcome markets, the outcomes are given optimism -1, -0.5, 0, 0.5, and 1; in two-outcome markets, they are given scores of -1 and 1). We limit the sample in Tables 5 and 6 to markets for which we can identify the outcome that would be good for the firm without making a difficult judgment call.

In Table 5, we find negative future returns for securities tied to optimistic outcomes in Google's markets. The evidence of the bias away from $1/N$ persists when controlling for optimism. There is also evidence of small biases towards purchasing, rather than selling securities (reflected in the negative average returns to expiry) and against purchasing securities tied to the most extreme outcomes (reflected in the positive returns for these securities). However, we do not see evidence of any of these biases in Firm X's markets. Likewise, the Ford sales markets do not exhibit

18. 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.

TABLE 5
Tests for pricing biases

Panel A: Google

	(1)	(2)	(3)	(4)
(Price – 1/N)		−0.232***	−0.226**	−0.210**
		(0.089)	(0.090)	(0.088)
Optimism			−0.103**	−0.106***
(+1 if best outcome, −1 if worst)			(0.041)	(0.040)
Extreme outcome				0.130**
Abs(Optimism)				(0.055)
Constant	−0.017***	−0.010**	−0.006	−0.014***
(Captures short aversion)	(0.004)	(0.004)	(0.004)	(0.004)
Trades	37,910	37,910	37,910	37,910
Securities	612	612	612	612
Markets	157	157	157	157
R^2	0.000	0.025	0.067	0.079

Panel B: Firm X

	(1)	(2)	(3)	(4)
(Price – 1/N)		0.026	0.017	0.020
		(0.050)	(0.050)	(0.051)
Optimism			0.021	0.021
(+1 if best outcome, −1 if worst)			(0.021)	(0.021)
Extreme outcome				0.033
Abs(Optimism)				(0.054)
Constant	−0.003	−0.003	−0.010	−0.010
(Captures short aversion)	(0.013)	(0.013)	(0.014)	(0.014)
Trades	8910	8910	8910	8910
Securities	1704	1704	1704	1704
Markets	945	945	945	945
R^2	0.000	0.000	0.002	0.002

Notes: Standard errors in parentheses. Each observation is a trade; the dependent variable is the percentage point return to expiry (*i.e.* expiry value − price). 1/N represents a naive prior, with N equal to the number of outcomes for the market ($N = 2$ for binary 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. The extreme outcome measure is the absolute value of an outcome's optimism, less the mean of this value across a market's securities. In the sample size data, a security refers to a unique security with a specific pay-off and a market refers to a group of securities with related pay-offs (*e.g.* a group of securities tracking mutually exclusive outcomes). Standard errors are heteroscedasticity robust and allow for clustering on market and calendar month. ***p <0.01, **p <0.05, *p<0.1.

evidence of a combined optimism and short aversion bias (based on the near-zero constant in Table 3, Panel C, Column 4).

In Table 6, we test for optimism separately for markets on different subjects. As in Table 5, we limit the sample to markets for which identifying the outcomes that are better for the firm can be done without difficult judgments calls. For both Google and Firm X, all "Fun" markets are excluded. For Google, high demand for products, timely completion of projects, and high product quality are all regarded as good for Google. Markets on external news were assigned optimism scores by a member of the company (who was not Cowgill) in cases where the assignments were regarded as uncontroversial (if there was any doubt about which outcome was better for Google, we did not assign optimism scores for that market).

Firm X is a largely U.S.-based basic materials and energy producer. We code macroeconomic outcomes associated with a strong economy (*e.g.* high GDP growth, low unemployment, high employment, high industrial production, and high equity prices) as good for the firm. Most

TABLE 6
Biases by subsample

Panel A: Google

	(1)	(2)	(3)	(4)	(5)
	All	Demand Forecasting	Project Completion	Product Quality	External News
Good outcome		High demand	On time	High quality	See text
Price – 1/N	−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	6898
Markets	157	51	38	22	42
R^2	0.067	0.024	0.211	0.207	0.104

Panel B: Firm X

	(1)	(2)	(3)	(4)	(5)
Good outcome	All	Politics GOP wins	Policy GOP policies	Stocks High values	Growth Rapid growth
Price - 1/N	0.017	−0.058	−0.081	−0.257	−0.013
	(0.050)	(0.221)	(0.086)	(0.158)	(0.125)
Optimism	0.021	0.163∗	0.034	0.001	0.049
	(0.021)	(0.090)	(0.120)	(0.086)	(0.032)
Constant	−0.010	−0.026	−0.004	0.028	−0.019
	(0.014)	(0.055)	(0.075)	(0.060)	(0.029)
Trades	8910	449	382	1309	2205
Markets	945	35	12	53	425
R^2	0.002	0.119	0.005	0.018	0.003

Panel B (continued): Firm X

	(6)	(7)	(8)	(9)	(10)	(11)
Good outcome	Jobs More jobs	Commodities Higher prices	Exchange Rates Weak dollar	Eurozone No crisis	Energy Higher prices	Inflation Faster inflation
Price - 1/N	−0.453	0.699***	0.129	0.452***	0.412***	0.034
	(0.278)	(0.123)	(0.090)	(0.090)	(0.116)	(0.122)
Optimism	0.175***	−0.019	0.012	−0.050	0.020	−0.095**
	(0.044)	(0.043)	(0.066)	(0.057)	(0.051)	(0.041)
Constant	0.043**	0.019	−0.059	0.002	−0.062**	−0.006
	(0.017)	(0.038)	(0.062)	(0.017)	(0.024)	(0.018)
Trades	657	290	462	166	492	1429
Markets	39	35	46	20	41	93
R^2	0.088	0.143	0.005	0.120	0.055	0.032

Notes: Standard errors in parentheses. Regressions identical to those in Table 5, Column 3 are presented for subsets of the Google and Firm X markets. Only markets for which optimism can be signed are included, and thus all "Fun" markets are excluded. See text for more details on the rationale applied in signing the optimism of different categories of outcome. Standard errors are heteroscedasticity robust and allow for clustering on market and calendar month. ***p <0.01, **p <0.05, *p<0.1.

markets Firm X ran on commodity prices were on commodities it produced or assisted in the production of, hence for these markets we code high commodity prices as good for Firm X. We exclude markets if we are uncertain about whether Firm X was a net buyer or seller of the commodity. Given the macroeconomic situation during the time period studied (2008–2013), we code increases in inflation as good for Firm X. During the 2011 European banking and sovereign debt crisis, Firm X ran markets on future interest spreads and write downs for investors, and we regard high spreads and write downs as negative for the global macroeconomy and thus for Firm X. For markets on exchange rates between the U.S. dollar and another currency, we code a weak dollar as good for Firm X, unless Firm X also produced in the country in question, in which case we omit that market from the sample. For markets on policy and politics, we code "pro-business" outcomes as good for Firm X, such as electoral victories by U.S. Republicans or U.K. Tories, or the passage of policies backed primarily by these parties.[19] Where applicable, our optimism codings are consistent with public statements by the firm's executives. We suspect that most readers will regard all of these judgements as uncontroversial; however, the impact of reversing or omitting any of them can be ascertained from the disaggregated results in Table 6.

In Table 6, we find that the optimistic bias is largest for markets on project completion. There are several reasons to expect the bias to be largest in these markets. First, these markets are on outcomes that are most under Google employees' control, and thus perhaps the most influenced by overconfidence about one's own or one's colleagues' ability. Secondly, strategic concerns for biased trading by insiders may be larger for these markets, given that outcomes are more under employees' control. Thirdly, information about project completion is presumably less dispersed throughout the organization than information about demand or external news, and discouraging entry by arbitrageurs and making the potentially biased views of project insiders more influential.

Unlike optimism, the degree of bias away from $1/N$ does not vary statistically significantly across the categories of Google's markets.[20] 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 $-$ $1/N$) and optimism variables.

As discussed above, the optimism in Google's markets could arise for either strategic or behavioural reasons. To help distinguish among the two, we conduct tests for company-wide "mood swings" in the optimism of 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 labour supply decisions, or good news for a company affecting future labour demand and thus hiring. Instead, we conclude that company-wide "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 5, Panel A, Column 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 1 standard deviation

19. This is consistent with the fact that stock market prices for basic materials and energy firms increase on average when Republicans win close elections (*e.g.* Snowberg *et al.*, 2007a,b and Zitzewitz, 2014).

20. The degree of optimism is statistically significantly different in the completion and external news categories (p <0.001 in both cases), but biases away from the prior are not statistically significantly different from one another ($p =$ 0.870). The p-values are calculated using versions of the regression in Table 6, Panel A, Column 1 that allow for add an interaction between the bias variable (*i.e.* price – $1/N$ or optimism) and an indicator variable for the market category.

TABLE 7
Optimism and stock returns

	(1)	(2)	(3)	(4)	(5)	(6)
Optimism*Google log stock return ($t+1$)	−0.869	−0.228	−0.303	−1.006	−0.646	−0.830
	(0.720)	(0.659)	(0.671)	(0.675)	(0.603)	(0.565)
Optimism*Google log stock return (t)	−1.158	−0.185	−0.243	−0.015	0.196	0.255
	(0.796)	(0.455)	(0.434)	(0.610)	(0.613)	(0.488)
Optimism*Google log stock return ($t-1$)	−2.022***	−1.318**	−1.296**	−2.618***	−2.112***	−1.414**
	(0.744)	(0.569)	(0.561)	(0.767)	(0.658)	(0.628)
Optimism*Google log stock return ($t-2$)	−0.695	0.037	0.063	−0.103	−0.043	−0.042
	(0.436)	(0.302)	(0.287)	(0.316)	(0.354)	(0.316)
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 price-1/N			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^2	0.095	0.155	0.159	0.489	0.505	0.522

Notes: Standard errors in parentheses. The regressions in this table extend the regression in Table 5, Panel A, Column 4 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), whereas columns 4–6 include only markets on the timing of project completion (*i.e.* those included in Table 6, Column 3). Standard errors are heteroscedasticity robust and allow for clustering within markets and calendar months. ***p<0.01, **p<0.05, *p<0.1.

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.

We conclude this subsection by examining how pricing biases evolve over time: over the life of an individual market and over the life of the prediction market experiment as a whole. As discussed above, the Firm X and Ford prediction markets use an automated market maker that is initialized at a prior, 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 8).[21] We find no evidence that prices are biased towards the price, even very earlier in a market's life.[22] 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 8 imply that they are actually biased away from the prior early in their life and that this bias abates with more trading history. As discussed above, the bias away from the $1/N$ prior suggests that traders are either overweighting their own prior beliefs or information that arrives after the market begins. The fact that the bias away from the market prior declines over the life of the market is more consistent with the former possibility. In contrast, the optimistic bias in Google's markets is small early in a market's life, and grows over time. This is consistent with market participants overreacting to new positive information and underreacting to new negative information.[23]

Finally, Table 9 presents tests of how the aforementioned (Price–Prior) and optimism biases evolved over our sample. Regressions from Tables 3 and 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 markets are large 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. Firm X's markets again appear efficient, albeit with weak evidence ($p = 0.09$) of a small optimism bias at the beginning of the sample that disappears by the sample's end.

2.4. *Individual trader characteristics and market efficiency*

This subsection analyses how traders' characteristics are 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.

21. 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 whether a trade is in a given decile of a particular market's life would not be known.

22. In an earlier version of the artcle, we cut the "Trades 1–10" sample even finer, finding no evidence of biases towards $1/N$, even in the prices of the first two trades in each market.

23. Unfortunately, we lack a direct measure of new information arrival for most of Google's markets. To further investigate over and under reaction, we ran tests for price momentum or reversals, but found that results were not robust to small changes in time horizons.

TABLE 8
Pricing biases over the life of markets

Panel A: Google	Trades 1–10	Trades 11–50	Trades 50+
(Price – Naïve prior)	−0.475***	−0.340***	−0.126
	(0.085)	(0.069)	(0.127)
Optimism	−0.013	−0.081**	−0.140**
(+1 if best outcome, −1 if worst)	(0.033)	(0.033)	(0.055)
Constant	−0.007*	−0.010*	−0.001
(Captures short aversion)	(0.004)	(0.005)	(0.007)
Trades	5,251	13,737	18,922
Markets	157	144	81
R^2	0.069	0.069	0.098

Panel B: Firm X	Trades 1–10	Trades 11–25	Trades 26+
(Price – Naïve prior)	0.003	0.055	−0.008
	(0.059)	(0.088)	(0.167)
Optimism	0.020	0.014	0.094
(+1 if best outcome, −1 if worst)	(0.019)	(0.042)	(0.102)
Constant	−0.006	−0.029	−0.038
(Captures short aversion)	(0.012)	(0.035)	(0.117)
Trades	7650	1,129	131
Markets	945	187	12
R^2	0.001	0.004	0.066

Panel C: Ford sales	Trades 1–10	Trades 11–50	Trades 51+
(Price – Naïve prior)	−0.122	−0.178	−0.811***
	(0.121)	(0.152)	(0.166)
Constant	−0.006	−0.011	−0.004
(Captures optimism and short aversion)	(0.008)	(0.008)	(0.005)
Trades	957	1747	558
Markets	101	86	20
R^2	0.034	0.059	0.710

Notes: Standard errors in parentheses. Regressions identical to those in Table 3, Column 4 are presented, except that trades in each security are numbered sequentially and the sample is split according to trade number. Standard errors are heteroscedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X). ***p<0.01, **p<0.05, *p <0.1.

To understand which traders contribute to and trade against pricing biases, we need to analyse the relationship between the nature of a position being taken (*e.g.* its optimism) and the characteristics of the trader. We begin by analysing all three companies, and thus focus on traders' past experience and past success. Before each trade, we calculate for each trader the number of prior trades that each trader has participated in and their average past return to expiry on all trades in contracts that have settled by that time. To be included in the sample, a trader must have at least one past trade in a contract that has settled.

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 by one trader and as a sell by another). 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 then regressed on

TABLE 9
Reduction in biases over time

	Google	Ford	Firm X
(Price – Prior)	−0.379***	−0.290**	0.063
	(0.134)	(0.109)	(0.120)
(Price – Prior)*Date	0.355	−1.251***	−0.114
	(0.287)	(0.227)	(0.212)
Optimism	−0.210***		0.088*
	(0.061)		(0.052)
Optimism*Date	0.274**		−0.129
	(0.115)		(0.082)
Constant	−0.010	−0.012	0.074
	(0.012)	(0.011)	(0.053)
Constant*Date (Min 0, Max 1)	−0.003	0.049*	−0.050
	(0.005)	(0.027)	(0.037)
Trades	37,910	3262	8910
R^2	0.090	0.26	0.006

Notes: Standard errors in parentheses. Regressions identical to those in Table 3, Column 4 (for Ford) and Table 5, Column 3 (for Google and Firm X) 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. Standard errors are heteroscedasticity robust and allow for clustering on market (for all three firms) and calendar month (for Google and Firm X). ***p <0.01, **p <0.05, *p<0.1.

trade fixed effects and the trader characteristics for that trade*side. This yields coefficients that are identical to what we would 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.[24] The coefficients in the regression tell us whether the traders with greater experience or better past returns is systematically on the purchasing side, on the optimistic side, on the side that buys securities priced above $1/N$ or sells those priced below, and on the side that ultimately earns positive returns.

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. Since we have one observation per trade rather than two, trade fixed effects are not included.[25] In these regressions, the coefficients tells us whether traders with greater experience or better past returns are more likely to buy than sell, are more likely to trade in an optimistic direction, are more likely to buy when prices are above $1/N$ and sell when they are below, and are more likely to buy securities that ultimately have positive returns to expiry.

Table 10 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 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.

24. 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.

25. We include time period fixed effects (for weeks for Ford and months for Firm X) to control for changes in trading behaviour over time, although doing so has limited impact on the results.

TABLE 10
Biases, experience and cumulative returns

Panel A: Google

	(1) Optimism	(2) Price – Prior	(3) Buy	(4) Returns
Cumulative returns	−0.520**	−0.036	−0.587**	0.178**
	(0.215)	(0.065)	(0.247)	(0.071)
Experience	−0.019**	−0.032***	−0.122***	0.015***
	(0.009)	(0.003)	(0.019)	(0.003)
Observations	75,820	141,412	141,412	141,412
R^2	0.005	0.049	0.055	0.006

Panel B: Ford sales

	(1) Buy/Optimism	(2) Price – Prior	(3) Returns
Cumulative returns	−0.149	−0.017**	0.023***
	(0.126)	(0.007)	(0.006)
Experience	−0.131***	−0.006***	0.006***
	(0.025)	(0.002)	(0.002)
Observations	2810	2810	2810
R^2	0.023	0.01	0.019

Panel C: Firm X

	(1) Optimism	(2) Price – Prior	(3) Buy	(4) Returns
Cumulative returns	−5.804	0.927	10.658	6.984***
	(5.075)	(1.607)	(9.324)	(2.490)
Experience	0.002	0.021***	−0.068**	−0.011***
	(0.011)	(0.005)	(0.033)	(0.004)
Observations	8696	12,318	12,318	12,318
R^2	0.001	0.018	0.010	0.005

Notes: Standard errors in parentheses. This table presents regressions testing whether traders with more past experience or higher past returns trade in a direction that is correlated with certain 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 data set twice. For Ford and Firm X, participants trade with an automated market maker, and hence 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 heteroscedasticity robust and allow for clustering within participants and markets. Regressions include fixed effects for trades for Google and time periods for Ford and Firm X (weeks and months, respectively). ***p <0.01, **p <0.05, *p<0.1.

In the Ford 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 both types of traders are more likely to sell when price is above its initial value (Panel B). The results presented in Sections 2.1 and 2.3 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 experience or 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 11
Trader characteristics, biases and returns at google

	(1) Optimism	(2) Price – Prior	(4) Buy	(5) Returns
Market insider	0.127	0.066*	0.401***	−0.012
	(0.099)	(0.035)	(0.153)	(0.046)
Friend of insider	0.146**	0.033*	−0.200	−0.026
	(0.069)	(0.019)	(0.179)	(0.019)
Coder/engineer	−0.008	−0.156***	−0.467***	0.102***
	(0.084)	(0.028)	(0.133)	(0.031)
Hire date (in years)	0.073**	−0.014	−0.152**	0.007
	(0.031)	(0.010)	(0.062)	(0.012)
NYC-based	−0.152	−0.117***	−0.089	0.064*
	(0.115)	(0.033)	(0.159)	(0.037)
Mountain-view based	−0.177*	−0.039	−0.130	0.037
	(0.094)	(0.026)	(0.105)	(0.025)
Observations	75,820	141,412	141,412	141,412
R^2	0.009	0.046	0.065	0.005

Notes: Standard errors in parentheses. This Table presents regressions analogous to those in Table 10, 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 heteroscedasticity robust and adjust for clustering within participants and markets. ***p<0.01, **p<0.05, *p<0.1.

In a previous version of the article, we analysed how continued participation by a particular trader was related to past performance and activity. At all three firms, continued participation more likely for traders with higher past returns and those who were more active in the prior period (Table A1 in the Online Appendix available as Supplementary Data). The reduction of pricing biases over time at Google and Ford are consistent with the fact that the more skilful and experienced traders trade against these biases, that traders gain experience over time, and that the most engaged and skilful traders are more likely to continue to participate.

Finally, we analyse the relationship between traders' job assignments and their prediction market trading using data that are only available to us for Google. Table 11 presents regressions with the same structure as in Table 10, 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 knowledgeable 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 knowledgeable employees trading in the opposite direction, pushing prices back to their efficient level.

It is also interesting that newly hired employees' 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 stretch targets rather than unbiased forecasts. Consistent with this, we find in unreported results 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, sales forecasts that are taken extremely seriously.[26] Secondly, 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 calibrated prediction markets being the markets at Firm X. The Firm X markets ran for almost 5 years and the average participant made over 200 trades.

Regarding the inefficiencies, some results match well with the prior literature, whereas 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 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" (*i.e.* that is it correlated with daily stock returns) is more consistent with optimism having at least a partly behavioural 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 our 2005–2007 sample period is also consistent with initial miscalibration and learning. Taken together, the evidence suggests that strategic biases, overconfidence, behaviour biases, and inexperience (*i.e.* beginning 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 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 lowpriced 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 their own priors or 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, and Firm X markets. We are limited to analysing the markets of firms who shared data with us, and the decision to share data may have been related to the success of the prediction markets. Nevertheless, we regard the evidence on the efficiency of corporate prediction markets as largely encouraging.

Producing efficient forecasts that improved upon the available alternatives was only one of the goals 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

26. We cannot distinguish whether it is the prediction market mechanism *per se* that leads to the better predictive performance, or simply the involvement of more people. It is possible that an averaging of forecasts from multiple experts, or a Delphi method approach to aggregating information from several forecasters would have also outperformed a single expert (or in Ford's case, a forecasting group). See Graefe and Armstrong (2011) for a laboratory experiment that compares the predictive performance of other group forecasting methods, such as Delphi.

having been embarrassed by prediction market trading into a redoubling of efforts, provides at least one example of this working.[27] We are aware of contrasting anecdotes from other companies though. For example, we are aware of four cases at different companies (outside those in our sample) 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).

Among our three sample firms, only Firm X's markets were still at the time we collected our data, despite relatively strong predictive performance at all three firms. The shutdown of Ford's sales markets is especially puzzling, as a 25% reduction in the mean-squared error of a sales forecast is presumably of significant value to an automaker. Our contacts at Ford tell us that budgets for experiments like prediction markets were limited by the still recessionary economic environment in 2010. Confidentiality concerns may have limited the usefulness of the features markets, and given that these markets accounted for the majority of trades, they may have overshadowed the more successful markets on sales. It is also possible that accurate sales forecasting is slightly less important during periods of overcapacity, like 2010.[28] Nevertheless, we still regard this decision as puzzling.

In Google's case, its prediction markets were begun as a "20% project" (Google allows it's engineering to spend up to 20% of their time working on a new project of their choice) in 2005. All members of the project team left full-time employment at Google around 2008–2010, and hence continuing the project would have either required recruiting new 20% engineers or assigning engineers to work on the project as their "80%" assignment. Engineers tend to prefer working on 20% projects of their own creation, and the bar for promotion of a 20% project is high.[29] One possible view of the non-continuation of Google's prediction markets is that opportunity cost of its engineers' time is high, and the "20% time" system intentionally sacrifices moderately successful projects to maximize the number of major successes.

An alternative view is that Google lacked the high-value application for prediction markets that Ford arguably had in sales forecasting. Forecasting demand for Google's products (such as Gmail) is probably less important than forecasting car sales as share of marginal costs are presumably a much lower share of total costs in Google's case, and acquiring processing and storage capacity in response to anticipated demand is a much more reversible decision than building a specific model of car, at least assuming that processing and storage capacity has many alternative applications. If running markets on project completion increased the likelihood that projects would be completed on time, this would presumably of much greater value to Google. Unfortunately, as mentioned above, Google's markets lacked the sample size needed to run an experiment designed to test for these effects. Furthermore, it is not obvious *a priori* that the effects would be positive. As discussed above, running markets on OKRs draws attention to them, presumably intensifying reputational incentives to achieve them. At the same time, if a project's participants hold initially

27. We originally hoped to produce more systematic evidence on this point by randomizing which OKRs were covered by markets, 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.

28. Ford sales in 2010 had rebounded about 20% from the low in 2009, but were still well below pre-2008 levels. See, *e.g*. Ford Motor Company, *2011 Annual Report*, 180. http://corporate.ford.com/doc/ 2011%20Ford%20Motor%20Company%20AR_LR.pdf (last accessed 27 June 2014).

29. Past 20% projects that have become "80% time" products include Gmail, Google News, Google Talk, suggested completions of Google queries, and AdSense (a second advertising system for Gmail and blogs that now accounts for 20% of its revenue). See, *e.g*. Tate (2013).

optimistic views of the likelihood that it will be completed on time, debiasing these views may not be in the company's interest.

Decisions about the adoption and continuation of corporate prediction markets are typically made by agents for organizations populated by other agents. Decisions about the adoption of corporate prediction markets may, therefore, depend on factors other than their utility in aggregating information. First, if agents in organizations earn rents from asymmetric information, adopting technologies that increase transparency may not be in their interests. A prediction market may provide an *ex ante* measure of a key project's expected quality, where otherwise only *ex post* measures (*e.g*. market acceptance) would have been available. Agents, including CEOs, may prefer noisier measures of performance, especially if performance is expected to be disappointing.

Alternatively, senior managers may have legitimate concerns about organizational side effects. Aggregating information about the success or failure of a key initiative helps informs management, but also informs other members of the organization. Informing these other members may have side effects, such as potentially adverse effects on effort levels, the leakage of information to competitors, or the facilitation of insider trading. The third concern limited the OKRs eligible for markets at Google, and the second concern constrained the design of Ford's markets on features. Informing an organization about a coming failure may be more damaging, in terms of morale, effort reduction, and employee turnover, than informing about coming successes is beneficial. That may be particularly true if employees have optimistic biases that benefit employers and that information sharing will reduce in expectation.

Our initial motivation for our analysis was that there were several plausible reasons to expect that prediction markets would not work well in corporate settings. Compared with public prediction markets, corporate markets are thinner, involve traders with potential biases, and have less potential for entry by arbitrageurs who reduce pricing biases. Despite this, the corporate prediction markets we study performed quite well. This, however, leaves us with two new open questions. First, why are corporate prediction markets not more popular, including at firms that have already experimented with them? And secondly, does this lack of popularity itself reflect agency problems? Would firms' owners benefit from insisting on their adoption?

Acknowledgments. The authors would like to thank Ford, Google, and Firm X for sharing the data used in this article 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, David Rothschild, Richard Schmalensee, Jesse Shapiro, Kathryn Shaw, Noam Yuchtman, our anonymous referees, and seminar participants at the AEA meetings, AMMA 2009, Berkeley, Chicago, Google, Erasmus, INSEAD, INFORMS, the Kaufmann Foundation, 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, at Firm X we thank the executive responsible for running the prediction markets, 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.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

AGGARWAL, R. K., and WU, G. (2006), "Stock Market Manipulations", *The Journal of Business*, **79**, 1915–1953.

ALI, M. M. (1977), "Probability and Utility Estimates for Racetrack Bettors", *The Journal of Political Economy*, **85**, 803–815.

ALLEN, F. and GALE, D. (1992), "Stock-Price Manipulation", *Review of Financial Studies*, **5**, 503–529.

BENABOU, R. and TIROLE, J. (2002), “Self-Confidence and Personal Motivation”, *Quarterly Journal of Economics*, **85**, 871–915.

BENABOU, R. and TIROLE, J. (2003), “Intrinsic and Extrinsic Motivation”, *The Review of Economic Studies*, **70**, 489–520.

BERG, J., FORSYTHE, R., NELSON, F. and RIETZ, T. (2008), “Results from a Dozen Years of Election Futures Markets Research”, *Handbook of Experimental Economic Results*, **1**, 486–515.

BERGMAN, N. K. and JENTER, D. (2007), “Employee Sentiment and Stock Option Compensation”, *Journal of Financial Economics*, **84**, 667–712.

BERNARDO, A. E. and WELCH, I. (2001), “On the Evolution of Overconfidence and Entrepreneurs”, *Journal of Economics & Management Strategy*, **10**, 301–330.

CHEN, K.-Y. and PLOTT, C. R. (2002), “Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem” (Working Paper No. 1131, California Institute of Technology Social Science).

COLES, P. A., LAKHANI, K. R. and MCAFEE, A. P. (2007), “Prediction Markets at Google” (Case 607-088, Harvard Business School).

COMPTE, O. and POSTLEWAITE, A. (2004), “Confidence-Enhanced Performance”, *American Economic Review*, 1536–1557.

COWGILL, B. and ZITZEWITZ, E. (2013), “Mood Swings at Work: Stock Price Movements, Effort and Decision Making” (Mimeo, UC-Berkeley).

COWGILL, B., WOLFERS, J. and ZITZEWITZ, E. (2009), “Using Prediction Markets to Track Information Flows: Evidence from Google” (Mimeo, UC-Berkeley).

FAIR, R. C. and SHILLER, R. J. (1989), “The Informational Content of Ex Ante Forecasts”, *The Review of Economics and Statistics*, 325–331.

FORSYTHE, R., NELSON, F., NEUMANN, G. R. and WRIGHT, J. (1992), “Anatomy of an Experimental Political Stock Market”, *The American Economic Review*, 1142–1161.

FOX, C. R. and CLEMEN, R. T. (2005), “Subjective Probability Assessment in Decision Analysis: Partition Dependence and Bias Toward the Ignorance Prior”, *Management Science*, **51**, 1417–1432.

GILLEN, B. J., PLOTT, C. R. and SHUM, M. (2013), “Inside Intel: Sales Forecasting using an Information Aggregation Mechanism” (Mimeo, Caltech).

GJERSTAD, S. (2005), “Risk Aversion, Beliefs, and Prediction Market Equilibrium” (Mimeo, University of Arizona).

GOEL, A. M. and THAKOR, A. V. (2008), “Overconfidence, CEO selection, and Corporate Governance”, *The Journal of Finance*, **63**, 2737–2784.

GOLDSTEIN, I. and GUEMBEL, A. (2008), “Manipulation and the Allocational Role of Prices”, *The Review of Economic Studies*, **75**, 133–164.

GRAEFE, A. and ARMSTRONG, J. S. (2011), “Comparing Face-to-Face Meetings, Nominal Groups, Delphi and Prediction Markets on an Estimation Task”, *International Journal of Forecasting*, **27**, 183–195.

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

GROSSMAN, S. J. and STIGLITZ, J. E. (1980), “On the Impossibility of Informationally Efficient Markets”, *The American Economic Review*, **70**, 393–408.

HANKINS, R. and LEE, A. (2011), “Crowd Sourcing and Prediction Markets”, *CHI*. **11**.

HANSEN, J., SCHMIDT, C. and STROBEL, M. (2004), “Manipulation in Political Stock Markets–Preconditions and Evidence”, *Applied Economics Letters*, **11**, 459–463.

HANSON, R. D. (2002), “Decision Markets”, *Entrepreneurial Economics: Bright Ideas from the Dismal Science*, 79–85.

HANSON, R. (2003), “Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation” (Mimeo, George Mason University).

HANSON, R., OPREA, R. and PORTER, D. (2006), “Information Aggregation and Manipulation in an Experimental Market”, *Journal of Economic Behavior & Organization*, **60**, 449–459.

HANSON, R. and OPREA, R. (2009), “A Manipulator Can Aid Prediction Market Accuracy”, *Economica*, **76**, 304–314.

HIRSHLEIFER, D., LOW, A. and TEOH, S. H. (2012), “Are Overconfident CEOs Better Innovators?”, *Journal of Finance*, **67**, 1457–1498.

HOFFMAN, M. and BURKS, S. (2013), “Training Contracts, Worker Overconfidence, and the Provision of Firm-Sponsored General Training” (Manuscript, University of Toronto).

JIAN, L. and SAMI, R. (2012), “Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution”, *Management Science*, **58**, 123–140.

KAHNEMAN, D. and LOVALLO, D. (1993), “Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking” *Management Science*, **39**, 17–31.

KAHNEMAN, D. and TVERSKY, A. (1979), “Prospect Theory: An Analysis of Decision Under Risk”, *Econometrica: Journal of the Econometric Society*, 263–291.

KEYNES, J. M. (1936), *The General Theory of Employment Interest and Money*, (London: MacMillan).

LARKIN, I. and LEIDER, S. (2012), “Incentive Schemes, Sorting, and Behavioral Biases of Employees: Experimental Evidence”, *American Economic Journal: Microeconomics*, **4**, 184–214.

LEVY, S. (2011), *In the Plex: How Google Thinks, Works, and Shapes Our Lives*. (New York: Simon and Schuster).

MALMENDIER, U. and TATE, G. (2008), “Superstar CEOs”, *The Quarterly Journal of Economics*, **124**, 1593–1638.

MANSKI, C. F. (2006), “Interpreting the Predictions of Prediction Markets”, *Economics Letters*, **91**, 425–429.

MONTGOMERY, T. A., PAUL, M. S., CAVARETTA, M. and MORAAL, P. (2013), "Experience from Hosting a Corporate Prediction Market: Benefits Beyond the Forecasts", KDD Conference Paper, 11 August 2013 (available at http://dl.acm.org/citation.cfm?id=2488212).

NEWMAN, A. L. (2012), "Manipulation in Political Prediction Markets", *The Journal of Business, Entrepreneurship & the Law*, **3**, 1.

OPREA, R., PORTER, D., HIBBERT, C., HANSON, R. and TILA, D. (2007), "Can Manipulators Mislead Market Observers" (Manuscript, University of California Santa Cruz).

ORTNER, G. (1998), "Forecasting markets—An industrial application", *Universität Wien*.

OTTAVIANI, M. and SORENSON, P. N. (2007), "Outcome Manipulation in Corporate Prediction Markets", *Journal of the European Economic Association*, **5**, 554–563.

OTTAVIANI, M. and SORENSEN, P. N. (2015), "Price Reaction to Information with Heterogeneous Beliefs and Wealth Effects: Underreaction, Momentum, and Reversal", *American Economic Review*, **105**, 1–34.

OYER, P. and SCHAEFER, S. (2005), "Why Do Some Firms Give Stock Options to All Employees?: An Empirical Examination of Alternative Theories", *Journal of Financial Economics*, **76**, 99–133.

PETERSEN, M. A. (2009), "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches", *Review of Financial Studies*, **22**, 435–480.

PRENDERGAST, C. (1993), "A Theory of 'Yes Men'", *The American Economic Review*, 757–770.

RHODE, P. W. and STRUMPF, K. S. (2004), "Historical Presidential Betting Markets", *The Journal of Economic Perspectives*, **18**, 127–141.

RHODE, P. W. and STRUMPF, K. S. (2006), "Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data" (Mimeo, University of Arizona).

RIGOBON, R. and SACK, B. (2005), "The Effects of War Risk on US Financial Markets", *Journal of banking & finance*, **29**, 1769–1789.

SNOWBERG, E., WOLFERS, J. and ZITZEWITZ, E. (2005), "Information (in)Efficiency in Prediction Markets", *Information Efficiency in Financial and Betting Markets*, 366.

SNOWBERG, E., WOLFERS, J. and ZITZEWITZ, E. (2007a), "Partisan Impacts on the Economy: Evidence from Prediction Markets and Close Elections", *The Quarterly Journal of Economics*, **122**, 807–829.

SNOWBERG, E., WOLFERS, J. and ZITZEWITZ, E. (2007b), "Party Influence in Congress and the Economy", *Quarterly Journal of Political Science*, **2**, 277–286.

SNOWBERG, E., WOLFERS, J. and ZITZEWITZ, E. (2013), "Prediction Markets for Economic Forecasting", *Handbook of Economic Forecasting*, **2A**, 657–684.

SONNEMANN, U., CAMERER, C., FOX, C. R. and LANGER, T. (2011), "Psychological Biases Affect Economic Market Prices" (Unpublished Manuscript, University of Münster, Germany).

SUROWIECKI, J. (2004), *The Wisdom of Crowds*. Random House LLC, 2005.

TATE, R. (2013), "Google Couldn't Kill 20 Percent Time Even if It Wanted To" *Wired*, 21 August 2013 (available at http://www.wired.com/2013/08/20-percent-time-will-never-die/, last accessed 23 April 2015).

TETLOCK, P. (2008), "Liquidity and Prediction Market Efficiency" (Unpublished Manuscript, Columbia University).

TZIRALIS, G. and TATSIOPOULOS, I. (2012), "Prediction Markets: An Extended Literature Review", *The journal of prediction markets*, **1**, 75–91.

WOLFERS, J. and ZITZEWITZ, E. (2004), *Prediction Markets*. No. w10504. National Bureau of Economic Research.

WOLFERS, J. and ZITZEWITZ, E. (2006a), "Interpreting Prediction Market Prices as Probabilities. (Working Paper No. 12200. National Bureau of Economic Research).

WOLFERS, J. and ZITZEWITZ, E. (2006b), "Five Open Questions About Prediction Markets" in Robert, H. and Paul, T. (eds) *Information Markets: A New Way of Making Decisions* (AEI Press).

WOLFERS, J. and ZITZEWITZ, E. (2009), "Using Markets to Inform Policy: The Case of the Iraq War", *Economica*, **76**, 225–250.

ZITZEWITZ, E. (2014), "Do Gas Prices Vote for the Right? Political Contributions Via Price Distortions" (Mimeo, Dartmouth College).