"Hidden Profiles" in Corporate Prediction Markets: The Impact of Public Information Precision and Social Interactions

Liangfei Qiu

Warrington College of Business, University of Florida,

Gainesville, FL 32611 U.S.A. {liangfeiqiu@ufl.edu}

Hsing Kenneth Cheng

Warrington College of Business, University of Florida,

Gainesville, FL 32611 U.S.A. {hkcheng@ufl.edu}

School of Information Management and Engineering, Shanghai University of Finance and Economics,

Shanghai, China

Jingchuan Pu

Warrington College of Business, University of Florida,

Gainesville, FL 32611 U.S.A. {jingchuan@ufl.edu}

“Hidden Profiles” in Corporate Prediction Markets: The Impact of Public Information Precision and Social Interactions

Abstract

Recently, large companies are experimenting with corporate prediction markets run among their employees. In the present study, we develop an analytical model to analyze the effects of information precision and social interactions on prediction market performance. We find that increased precision of public information is not always beneficial to the prediction market accuracy because of the “hidden profiles” effect: the information-aggregation mechanism places a larger-than-efficient weight on existing public information. We show that a socially embedded prediction market with information sharing among participants may help correct such inefficiency and improve the prediction market performance. We also identify conditions under which increased precision of public information is detrimental in a non-networked prediction market and in a socially embedded prediction market. These results should be of interest to practitioners as the managerial implications highlight the detrimental effect of public information and the role of social networking among employees in a corporate prediction market.

Keywords: prediction markets, social networks, public information, information sharing, hidden profiles

"Prediction markets comprising a diverse set of consumers can be valuable tools for companies spanning a wide range of industries and at every stage in the product or service life cycle and be used for: narrowing the new-product development funnel; concept testing; forecasting; pricing; message optimization; and promotion testing."

— Julie Schlack, Senior Vice President at Communispace[1]

"Actors do not behave or decide as atoms outside a social context."

— Granovetter (1985), p. 487

1 Introduction

How to take advantage of the knowledge dispersed in various parts of their organizations for better decision making has been a persistent challenge for large corporations. Lew Platt, the former chief executive of Hewlett-Packard, observed that, "if only HP knew what HP knows, we would be three times more productive."[2] Enterprise collective intelligence isn't about changing a specific industry, but rather revolutionizing the nature of how businesses operate and changing the landscape of corporate decision making. In the classical command-and-control model of businesses, corporate decision making is generally hierarchical. The manager of each business department is responsible for acquiring all information from that department, synthesizing it, and making decisions or reporting it up the chain of command. While this approach can sometimes lead to an optimal quantity and quality of information reaching the corporate decision maker, it also has a significant informational problem: The information necessary to make a decision can be filtered or distorted in the hierarchy (Abaramowicz and Henderson 2007).

An alternative model of decision making is to facilitate the flow of information around the hierarchy by tapping into and exploiting the collective intelligence within the organization. A common

[1] See http://www.quirks.com/articles/2013/20130509.aspx.

[2] See http://www.bloomberg.com/bw/stories/2008-12-22/harrahs-new-twist-on-prediction-marketsbusinessweek-business-news-stock-market-and-financial-advice.

application of corporate collective intelligence is a prediction market—a speculative or betting market that invites participants to speculate on uncertain future events. It works similarly to a financial market. A risky "asset" is defined as reflecting an issue of interest to the company such as the product's readiness to launch and sales trends of a new product. Prediction market prices have informational value because they aggregate the beliefs of market participants and reveal what the market's overall forecasts are.

Companies have made increasing use of prediction markets to help make business decisions. As documented in Chen and Plott (2002), Cowgill, Wolfers, and Zitzewitz (2009), and Cowgill and Zitzewitz (2014), a number of companies in a broad range of industries, such as Hewlett-Packard, Intel, BestBuy, Microsoft, Ford, Chrysler, Google, Eli Lilly, General Electric, and Siemens, have begun experimenting with corporate prediction markets. The prediction tasks vary from drug development success at Eli Lilly[3] to monthly operating profits and revenues at Hewlett-Packard (Chen and Plott 2002); from a project completion date at Siemens (Leigh and Wolfers 2007) to allocation of manufacturing capacity at Intel (Hopman 2007); and from weekly vehicle sales at Ford (Montgomery et al. 2013) to the number of Gmail users at Google (Cowgill, Wolfers, and Zitzewitz, 2009). Early evidence on the performance of corporate prediction markets has been encouraging: Similar successes have been repeatedly observed in many corporate prediction markets, and the prediction markets outperform existing mechanisms in terms of forecasting precision.

A fundamental innovation of corporate prediction markets is to introduce a collaborative market mechanism to augment the hierarchical decision-making process and improve overall decision quality. Essentially, a prediction market is a smart market system that can provide decision support in complex environments (Bichler, Gupta, and Ketter 2010). Coase (1937) explained that the existence of firms is driven by the fact that the benefits of hierarchy and command-and-control exceed the transaction costs. Prediction markets have the potential to profoundly reduce the costs of hierarchy by allowing

[3] https://www.crowdworx.com/news/crowdworx-interview-carol-gebert-prediction-markets-pharmaceuticals-industry/.

information to flow to top decision makers (Abaramowicz and Henderson 2007). CEOs and high-ranking managers are tasked with filtering and analyzing information; however, the data that they receive is also filtered (and distorted) by their subordinates. The social pressures within a company may lead some employees to conform to existing public information, such as official reports or the opinion of some high-ranking managers, although these employees may have valuable information to share—an effect known as "hidden profiles" (Stasser and Titus 1985; Stasser and Titus 2003; Sunstein 2005). The original idea of the hidden profile effect describes a biased pattern of information distribution in which some information, prior to group discussion, is shared by all group members (public information), and some is unique to individual members (private information). Group members often fail to effectively pool their information because discussions tend to be dominated by public information that members hold before the discussion. In other words, hidden profile refers to the phenomenon of overweighting public information in general, and it can arise from different sources. Prendergast (1993) proposed an economic theory of "yes men," where employees have an economic incentive to conform to the opinion of their supervisors. Therefore, one potential source of hidden profile is that individual employees tend to place a larger weight on *existing public information* than justified by its informational content when they report their opinions to upper level decision makers in corporate hierarchy.

Prior literature argued that a prediction market can help alleviate the hidden profile effect—the overreaction of corporate decision making to existing public information (Abaramowicz and Henderson 2007). The basic logic is as follows: In a corporate prediction market, employees would have incentives to correct the conformity and overreaction to public information, especially if they can trade anonymously. As highlighted by Bo Cowgill, a Google economic analyst, the anonymous trading system in prediction markets lets the Google hierarchy discover its employees' uncensored opinions. "If you let people bet on things anonymously, they will tell you what they really believe because they have money at stake. This is a conversation that's happening without politics. Nobody knows who each other is, and nobody has any incentive to kiss up."[4] Consequently, the use of corporate prediction markets can help decision makers reduce the possibility that errors propagate through the hierarchy all the way to the top, as individual employees who dissent from an official report would have a financial incentive to trade against the official report. The market approach with the protection of anonymity gives a voice to internal employees who otherwise would be affected by hidden profiles or yes men due to various pressures or expected costs from speaking out.

However, can a corporate prediction market completely solve the problem of overweighting public information? In this paper, we develop an analytical model of market trading to analyze the impact of information quality on prediction market performance. As defined in prior literature (Morris and Shin 2002; Chen and Jiang 2006; Angeletos and Pavan 2007), we differentiate between two types of information within an organization according to the way the information is generated: (i) public information that is common to all prediction market participants, such as official company reports, and (ii) private information that can be accessed only by individual employees, such as tacit knowledge from their working experience. Although prediction market participants place individually optimal Bayesian weights on both public and private information in our model, we find that a prediction market can cause another type of hidden profile effect: The information-aggregation mechanism in a corporate prediction market will place a larger (less) than efficient weight on public (private) information. This result suggests that even if the effect of hidden profiles at the individual level can be corrected by a corporate prediction market, the information-aggregation mechanism (market mechanism) can be another source of the "hidden profile" effect (the problem of overweighting public information).

A key assumption of our model is that a prediction market participant lacks the ability to extract other participants' information from market prices. Because of this bounded rationality, a corporate prediction market will place a larger-than-efficient weight on public information. Essentially, our model

[4] See http://www.networkworld.com/article/2284098/data-center/google-bets-on-value-of-prediction-markets.html.

is different from the efficient market literature (Malkiel and Fama 1970) in how a competitive market serves to communicate information between the market participants. The classic efficient market assumption states that the equilibrium price aggregates all the available information in the market perfectly, and participants have unlimited cognitive abilities to process information (Malkiel and Fama 1970; Grossman 1976, 1978). However, if the equilibrium price really aggregates all the available information perfectly as Grossman (1976, 1978) suggested, participants will neglect their own private information because it is useless. According to this logic, it is unclear why the price should reflect the private information in the first place (Hellwig 1980). A number of empirical studies in finance also show that a competitive market does not aggregate information as efficiently as we expect (Jensen 1978; Shiller 1981; Bondt and Thaler 1985). Our study relaxes the assumption that the equilibrium price aggregates information perfectly by introducing more realistic constraints on participants' information-processing abilities.

Another key finding of our research is that increased precision of private information always enhances prediction market performance as expected, but, surprisingly, increased precision of public information is detrimental to the prediction market performance when public information is relatively noisy. An intuitive explanation is that the presence of public information might have a distortive effect on the prediction market price formation. All prediction market participants receive the same public information. As such, each participant will form her best guess according to her own private information as well as the same public information. In the process of aggregating all participants' best guesses, a corporate prediction market mechanism will count the public information multiple times. Therefore, the aggregated prediction market forecast will overreact to the public information, and any noise contained in the public information will be magnified.

More importantly, we uncover the specific mechanism through which the problem of overweighting the public information might be mitigated, which is an issue unaddressed by the extant literature. In particular, social interactions and information sharing among prediction market participants may help correct the overreaction to public information. Actually, a distinct feature of a corporate prediction market versus a public prediction market, such as Iowa Electronic Markets (Berg et al. 2008), is that internal employees are more likely to be socially connected: They can exchange information with each other through personal networks and social relations, which are important conduits of knowledge (Qiu, Rui, and Whinston 2014a, b).[5]

As social media technologies have grown explosively, employees are increasingly using public social media platforms such as Twitter, Facebook, and LinkedIn for work-related purposes (Parise, Whelan, and Todd 2015). Many companies, such as 7-Eleven, Capital One, and Dow Chemical, have developed their own in-house corporate platforms to promote social networking among employees (Mello 2014). Knowledge sharing in knowledge networks becomes much more common in today's digitally connected world (Hansen 2002). Therefore, beyond a non-networked corporate prediction market with independent bettors, we examine a model of a socially embedded prediction market, in which participants can share information with their social connections.

In our study, we highlight that public information is a double-edged instrument in a prediction market. It conveys information on the fundamentals of the asset traded in the prediction market, but on the other hand, the noisiness of public information can be enhanced in the prediction market due to the overreaction to the disclosure of public information. Corporate managers should be aware that increased precision of public information might have a detrimental effect on the aggregation of information into prediction market prices. Our results on socially embedded prediction markets further illustrate the complex interaction between private and public information.

2 Literature Review

The literature on prediction market design has been growing rapidly in recent decades (e.g. Guo, Fang, and Whinston 2006; Fang, Stinchcombe, and Whinston 2007, 2010; Berg, Neumann, and Rietz

[5] As a philosophical matter, Google's rule for managing knowledge workers is to pack people in tight, so they can share information (Cowgill, Wolfers, and Zitzewitz 2009).

2009; Healy et al. 2010; Van Bruggen et al. 2010; Jian and Sami 2012; Cowgill and Zitzewitz 2014). Most of these studies, explicitly or implicitly, assumed that prediction market participants are isolated in the sense that they cannot communicate their private information with each other. As we have stated, in a corporate internal prediction market, participants are more likely to be socially connected. Although a handful of recent empirical studies have begun considering socially embedded prediction markets (Cowgill, Wolfers, and Zitzewitz 2009; Qiu, Rui, and Whinston 2014a), analytical analysis is limited in this area. To bridge this research gap, our work provides a modeling framework to understand the role of information precision in a socially embedded prediction market, and to improve the design of corporate prediction markets.[6]

Our work is more broadly related to the literature on the wisdom of crowds. An intriguing question is about the boundary conditions of crowd wisdom: When is a crowd wise? In an analytical model, Golub and Jackson (2010) showed that whether a crowd is wise depends critically on the structure of social networks. Lorenz et al. (2011) studied a forecast-report context in which the crowd prediction is a linear combination of all individuals' predictions, and found experimental evidence that sharing information among individuals may undermine the wisdom of crowd effect. In a similar context, Davis-Stober et al. (2015) demonstrated how to create optimal forecasting groups. In the context of expert-systems design, Jiang, Mookerjee, and Sarkar (2005) considered a sequential information gathering problem in which input data may be distorted by system users. The optimal design of other smart systems, such as internal knowledge investment (Ba, Stallaert, and Whinston 2001), consumer contests (Liu, Geng, and Whinston 2007), consumer review systems (Jiang and Guo 2015) and dynamic electricity trading systems (Ketter et al. 2015, 2016), has also been widely examined in the literature. Our model focuses on a specific form of the wisdom of crowds—a prediction market in which the forecast is generated from

[6] Jian and Sami (2012) differentiated two commonly used mechanisms of prediction markets—probability-report mechanism and security-trading mechanism. Qiu, Rui, and Whinston (2014a, b) mainly focused on the probability-report mechanism in a socially embedded prediction market using controlled laboratory experiments. The focus of our study is to derive analytical insights on the security-trading mechanism in a socially embedded prediction market, but we also show that our results are robust in a forecast-report mechanism in online appendix E.

market prices using a security-trading mechanism.

Recent empirical studies in finance documented evidence that social networks play an important role in financial markets (Coval and Moskowitz 2001; Cohen, Frazzini, and Malloy 2008). A stream of analytical studies showed that social communications among traders improve market efficiency (Colla and Mele 2010; Ozsoylev and Walden 2011; Han and Yang 2013). Our work differs from theirs for two reasons. First, in our study, we focus on the role of public information in a non-networked prediction market versus a socially embedded prediction market. Although Han and Yang (2013) pointed out that increased precision of private information improves price informativeness, none of these studies have considered the detrimental effect of public information. Our analytical results complement their research and highlight that the effect of information precision on market efficiency depends on the way the information is generated. Increased precision of private information always enhances prediction market performance regardless of whether a social network is embedded, but increased precision of public information could be detrimental under some market conditions. Second, this stream of studies adopted the large economy analysis by assuming the number of market participants is infinity, and investigated the asymptotic properties of an equilibrium. This approach is well defined and valid in financial markets. However, unlike a financial market or a public prediction market, a thin market is an important feature of a corporate prediction market because of confidentiality reasons. There is often a need to limit participation for prediction topics with strategic importance (Cowgill and Zitzewitz 2014). Therefore, our model focuses on the case that the number of prediction market participants is limited. We demonstrate that the number of prediction market participants has a significant impact on when increased precision of public information is more likely to be detrimental.[7]

Following the prior finance literature on difference of opinions (DO), in our model, we assume that prediction market participants do not condition on prices to infer private information of others. The

[7] When the number of participants is large, our results can be applied to public prediction markets.

DO behavior can be explained by (i) behavioral biases such as bounded rationality and limited attention, or by (ii) heterogeneous priors among rational agents (Banerjee, Kaniel, and Kremer 2009; Banerjee and Kremer 2010). From a perspective of behavioral biases, Hong and Stein (1999) proposed a DO model to explain the momentum phenomenon in financial markets. Hong and Stein (2003) developed an analytical framework of market crashes based on difference of opinions among investors.[8] Scheinkman and Xiong (2003) used overconfidence as a source of difference of opinions to examine speculative bubbles in asset prices. Banerjee, Kaniel, and Kremer (2009) showed that difference of opinions is necessary to generate price drift, which is an empirical regularity in financial markets.

From a perspective of heterogeneous priors, early studies used the DO approach to avoid the No Trade theorem resulting from rational expectations, and to generate positive trading volume in analytical models (Harrison and Kreps 1978; Harris and Raviv 1993; Kandel and Pearson 1995). They demonstrated that DO can generate trading patterns consistent with stylized empirical evidence. Cao and Ou-Yang (2009) analyzed the effects of DO on the dynamics of trading volume in stocks. Banerjee and Kremer (2010) developed a dynamic DO model to generate positive autocorrelation in trading volume.

Our paper differs from the prior DO studies in several aspects. First, the previous DO literature has mainly focused on using DO to explain a number of empirical features of price and volume dynamics in financial markets, such as momentum and positive autocorrelation in trading volume. However, our paper focuses on prediction market accuracy—more specifically, the impact of public information on prediction market accuracy in the DO framework. Second, a few studies have noticed that the problem of overweighting public information can be caused by the DO framework, and the overweight issue is the underlying driving force of some empirical regularities in financial markets, such as price drift (e.g. Banerjee, Kaniel, and Kremer 2009). However, as far as we know, none of these studies have examined the following key results on the optimal design of prediction markets in our work: (i) increased precision

[8] In their model, each investor receives a private signal, and each investor's signal contains some useful information. However, each cognitively overloaded investor only pays attention to her own signal, even if other investors' signals are revealed in prices.

of public information is not always beneficial to prediction market accuracy and can be detrimental when public information is relatively noisy; (ii) social networks among employees help correct the problem of overweighting public information and improve prediction market accuracy; (iii) the social network, however, has a side effect. As the level of social interaction increases, increased precision of public information may be more likely to be detrimental under some market conditions; and (iv) although social interactions can correct the overweighting problem, the homophily effect (the errors of friends' private signals are positively correlated) tends to weaken the correction because friends' information is less useful under homophily.

Our study is also related to social psychology literature on hidden profiles. The implication from the prior psychology experiments is that individuals tend to attach a larger weight to shared common information than justified by its informational content in their decision-making process (Stasser and Titus 1985; Stasser and Titus 2003). A stream of economics literature on the role of public information (Morris and Shin 2002; Angeletos and Pavan 2007) found a similar result when individuals have coordination motives using analytical models.[9] Instead of focusing on the effect of hidden profiles in the individual level, our model digs deeper into the overweighting problem of the information-aggregation mechanism (the effect of hidden profiles) in prediction markets: Even if there is no distortion in the individual level, the problem of overweighting the public information can be caused by the prediction market mechanism itself. In our model, the social value of network communications is to alleviate the overreaction of prediction market prices to the public information.

3 A Benchmark Non-Networked Prediction Market

[9] The underlying mechanism that leads to the overweight issues in our paper is different from that in the prior literature. It is important to note that the context of Morris and Shin (2002) is a stylized coordination game: In each individual's utility function, there is a "beauty contest" term. It means that each individual's utility depends on the average level of actions of others. Essentially, this beauty contest term is the driving force of the overweight issues. Our model is a market setup, which is different from Morris and Shin (2002). More importantly, there is no "beauty contest" term in the utility function of prediction market participants. The driving force of the overweight issues is that the information-aggregation mechanism in prediction markets places a larger than efficient weight on public information. As far as we know, very few studies have looked at the overweight issues caused by the market aggregation mechanism, and our paper is the first to examine the inefficient multiple counting of public information in a market context (prediction market). Additionally, the prior literature on the overweight issues of public information (e.g. Morris and Shin 2002) has mainly focused on social welfare from a social planner's perspective. Increased precision of public information may be detrimental to social welfare. In this paper, we focus on the impact of public information on the overall prediction market performance instead of the total welfare of prediction market participants.

3.1 Model Setup

In a corporate prediction market, assets are created whose final value is tied to a particular event—for example, the sale of a new product. People trade the assets according to their forecasts. More specifically, in our model, people trade a single asset according to the outcome of a future random variable, V. A manager wants to forecast V, and she resorts to n prediction market participants (internal employees) to obtain an accurate prediction. Table 1 provides a list of the notations used in our model.

All the prediction market participants share a common prior on V, given by:

$$V \sim N(V_0, 1/\rho_V),$$

where V_0 is the mean of the prior, and ρ_V is the precision of the prior. Before the prediction market opens, each participant can access a private signal:

$$S_i = V + \varepsilon_i, \varepsilon_i \sim N(0,1/\rho_\varepsilon), \varepsilon_i \perp \varepsilon_j, \tag{1}$$

where ρ_ε is the precision of participant i's information source for $i = 1,2,\ldots,n$.[10] The signals' errors $\varepsilon_1,\ldots,\varepsilon_n$ are independent across participants and are also independent of V. In this non-networked benchmark model, we assume that each participant accesses a private independent signal and does not communicate with each other.[11] Then, in later sections, we relax this assumption, and allow communication among friends and correlated private signal errors (homophily).[12]

Table 1. Summary of Notations

Notation	Description

[10] We assume that the precisions of all participants' private information are equal. It implies that no one is especially well informed, and that the valuable information is not concentrated in a very few hands. Our result is robust when we consider two groups of participants with heterogeneous precisions of private information (experts and ordinary participants). The simulation analysis can be found in online Appendix D. We also assume that the acquisition of private information is costless. Qiu et al. (2014a) investigated a model of costly information acquisition in prediction markets and found that the more friends a participant has, the less willing she is to acquire information.

[11] Our non-networked benchmark model is more like a public prediction market that has been widely studied in the literature. In public prediction markets (Berg and Rietz 2003; Berg et al. 2008, Foutz and Jank 2010), participants are assumed to be isolated: They receive bits and pieces of independent information and cannot communicate with each other. The reason is that in public prediction markets, participants are anonymous traders, and typically they don't know each other.

[12] It is worth noting that a rising share of employees now regularly engages in working from home, especially in the tech industry (Bloom et al. 2015). A byproduct of working from home is that the adoption of working from home may significantly reduce the intensity of social interactions among employees, and hence weaken communication within an organization. In our specific context, the rising practice of working from home may make corporate prediction market participants much more isolated than before. If a significant portion of prediction market participants works from home, employees are less likely to discuss and share information during coffee break or lunch time. Empirical evidence shows that the share of managers in the United States, the United Kingdom, and Germany allowed to work from home during normal hours is almost 50%. Working from home is also becoming increasingly common in developing countries because of rising traffic congestion and the spread of laptops and cell phone connectivity (Bloom et al. 2015).

V	A random future event that will be forecasted in a corporate prediction market
S_i	Each individual's private signal
ε_i	The noise contained in the private signal
V_0	The mean of the common prior (public information)
ρ_V	The precision of public information
ρ_ε	The precision of private information
n	The number of prediction market participants
x_i	Each individual's trading position in a prediction market
π_i	Each individual's trading profits
I_i	Each individual's information set
k	The number of friends each participant has in a regular network
a_0	The proportion of degree 0 participants
$N_i(g)$	The set of individual i's friends
γ	Risk averse parameter
δ	Correlation coefficient under homophily
m	The number of participants that use the DO approach to make inferences

We provide a running example of corporate prediction markets to show how our model setups are tied to reality, as follows: In the business practice, the random variable V could refer to the next month's sales of a product or the following quarter's monthly sales of a product (Chen and Plott 2002). In Google or Ford, there is a prediction market associated with every event they are trying to predict, such as the growth rate of Gmail users in Google (Cowgill, Wolfers, and Zitzewitz, 2009) or sales volumes for selected Ford models (Montgomery et al. 2013).

The common prior can be interpreted as the existing public information available to prediction market participants. Many projects at Google had "dashboards," or online summaries of project status. These dashboards are typically visible to all Google employees and can be treated as public information. If Google contains a prediction market related to a project that has a dashboard, the prediction market webpage includes a link to the dashboard (Coles, Lakhani, and McAfee 2007). Actually, many high-tech companies use dashboards to track project status. Private signals reflect the diverse information that can be assessed by internal employees in their daily work. For instance, if data analysts in a company's marketing department are asked to predict future sales of a product, they may have different information sources from their working experience.[13] The private signal in Google's prediction markets refers to

[13] The focus of this study is to examine the impact of information precision on prediction market performance. Therefore, we follow the prior literature

information only available to a small number of individuals or personal interpretations. For instance, a manager on the search quality team in Google mentioned that she had private information when a prediction market was related to her own projects. Another Google prediction market participant said: "the one time I thought I had good (private) information on a Google project, where there was a market I traded like crazy on it." (Coles, Lakhani, and McAfee 2007, page 12).

In a competitive security-trading prediction market, participants trade anonymously, taking prices as given. Participant i's profits are given by $\pi_i = (V - P)x_i$, where P is the prediction market price of the risky asset tied to V, and x_i is the demand for the security of participant i. If $x_i > 0$, participant i holds a positive position in the risky asset; if $x_i < 0$, participant i shorts the risky asset. We further assume that participant i's preferences over random profits are described by a mean-variance utility function, which has been widely adopted in the economics and finance literature (Levy and Markowitz 1979; Aid et al. 2011). Therefore, participant i is risk averse, and her utility depends on the expected profits as well as the variance of the random profits:

$$\mathbf{E}[\pi_i] - \gamma \mathbf{Var}[\pi_i], \tag{2}$$

where γ is a parameter that captures the risk aversion of participants. If γ is larger, participants are more risk averse. Following the finance literature (e.g. Cespa and Vives 2015), we assume all participants share the same risk aversion parameter γ to simplify the calculation.

Our benchmark model is a two-date static model. Following the standard timeline setup in the finance literature (Grossman and Stiglitz 1980; O'hara 1995; Banerjee, Kaniel, and Kremer 2009; Cespa and Vives 2015), we describe the timeline of our model as follows. At time 1, prediction market participants receive private signals and trade the risky asset. The market price of the asset and the trading position of each individual are simultaneously determined. At time 2, the random variable V is realized.[14]

(Morris and Shin 2002; Chen and Jiang 2006; Angeletos and Pavan 2007) and assume that both private and public signals are unbiased.

[14] Following the finance literature (Grossman and Stiglitz 1980), we use the way modeling financial markets to model prediction markets. The reason is

According to equation 2, participant i's optimization problem of choosing the optimal quantity becomes:

$$\max_{x_i} \mathbf{E}[(V-P)x_i|I_i] - \gamma \mathbf{Var}[(V-P)x_i|I_i], \quad (3)$$

where I_i is the information set of participant i. The first order condition (F.O.C) yields:

$$x_i^* = \frac{\mathbf{E}[V|I_i]-P}{2\gamma \mathbf{Var}[V|I_i]}. \quad (4)$$

In the prior finance literature, there are two major paradigms for modeling inference process in financial markets (what the information set I_i should contain): *rational expectation equilibrium* (REE) and DO. Both approaches share the view that investors have different valuations, and prices aggregate the different views during the trading process. The difference between the REE and DO models is in the information set of participants, I_i. "In an REE, an agent conditions both on the private signal and the price vector. In the DO model, however … each agent conditions only on his or her private signal" (Banerjee, Kaniel, and Kremer 2009, pp. 3712). More specifically, the REE approach implies that participants are able to learn from the price: the quantity they demand for a given price depends on the information the price reveals about the value of the asset (Grossman and Stiglitz 1980; O'hara 1995; Cespa and Vives 2015). The information set under REE should include the private signal as well as the expected price, $I_i = \{S_i, P^*\}$. In contrast, the DO approach assumes that participants are not as sophisticated as REE participants, and they do not learn from the information contained in market price. The information set under DO includes only the private signal, $I_i = \{S_i\}$.

In Section 3.2, we look at a pure DO model where all participants do not learn from the information contained in market price. The reality of corporate prediction markets is likely to be neither as efficient as in an REE nor as inefficient as in a pure DO equilibrium, but somewhere in between. Therefore, in Section 3.3, we develop a mixture model that nests both the REE and DO approaches:

that most successful prediction markets, such as prediction market in Google and Ford (Coles, Lakhani, and McAfee 2007; Montgomery et al. 2013), are very similar to stock markets: "they contained securities, each of which had a price. People used the market to trade with one another by buying and selling these securities (Coles, Lakhani, and McAfee 2007, p. 1).

among total n participants, m participants use the DO approach to make inferences (they do not learn from the expected price), and $n – m$ participants use the REE approach to make inferences (they are sophisticated traders and learn from the expected price). This modeling setup captures the heterogeneity of participants in cognitive capabilities. If $m = n$, the model converts to a pure DO model; if $m = 0$, the model converts to a pure REE model.

3.2 A DO Model of a Non-Networked Prediction Market

We first look at a pure DO model of prediction markets for two reasons. First, as documented in the prior literature (Banerjee, Kaniel, and Kremer 2009), the appeal of DO models is that the predictions from DO models are consistent with real-world empirical evidence in financial markets. For instance, a typical feature of REE in financial markets is the No Trade theorem (Tirole 1982): no trader expects a positive monetary gain from his trade; thus, the trading volume is zero. The intuition of this theorem is that in the REE framework (traders are completely rational), if one participant has information that induces her to want to trade at the current asset price, then other rational participants would be unwilling to trade with her, because they realize that she must have superior information. The no-trade result derived from the REE framework is not consistent with the empirical evidence observed in real-world corporate prediction markets (Montgomery et al. 2013; Cowgill and Zitzewitz 2014): Participants actively trade in corporate prediction markets. In contrast, the DO models do not generate the no-trade result and are consistent with the observed evidence. The empirical literature in financial markets (Banerjee and Kremer 2010) has documented that a number of regularities on observed levels and patterns of trading volume are difficult to reconcile in standard REE models (even in noisy REE models). In contrast, the DO framework appears better suited to address the empirical evidence involving trading volume (Banerjee and Kremer 2010). More broadly, Lovell (1986) summarized a number of empirical studies challenging the validity of rational expectation hypothesis in contexts other than financial markets.

Second, prediction market participants in our context are corporate employees, not professional traders in financial markets. For instance, an active trader in Google's prediction markets admitted: "I never play the real-world stock market." (Coles, Lakhani, and McAfee 2007, p. 12). Therefore, they may not be as sophisticated as professional traders who can learn from the information contained in the expected price. Essentially, REE requires extraordinary analytical and computational capabilities for a fully rational approach. It is highly impractical and cognitively demanding for non-professional traders in prediction markets to learn from the information contained in the expected price and to do the required calculations. The DO paradigm can be motivated by behavioral biases, such as bounded rationality (Banerjee and Kremer 2010), and this approach is more appropriate in our corporate prediction market context.

In a pure DO model, each participant makes a Bayesian inference using her private signal and the common prior:

$$\mathbf{E}[V|I_i] = \mathbf{E}[V|S_i] = \frac{\rho_V}{\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}S_i,\ \mathbf{Var}[V|I_i] = 1/(\rho_\varepsilon + \rho_V).$$

Essentially, participant i's conditional expectation, $\mathbf{E}[V|I_i]$, is a weighted average of the prior mean and her private signal. Note that $\frac{\rho_V}{\rho_\varepsilon+\rho_V}$ is the Bayesian weight given to public information, and $\frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}$ is the weight given to private information. We close the model by imposing the market clearing condition, which determines the prediction market price P: $\sum_{i=1}^{n} x_i^* = 0$, where n is the number of participants in the prediction market. The market clearing condition simply means that the sum of each participant's position should be equal to zero. We assume n is a limited number. However, our model can be easily extend to the case of $n \to \infty$.

We characterize the benchmark equilibrium without social networks when participants are allowed to trade assets in a competitive prediction market in the following proposition. All the proofs can be found in online Appendix A.

Proposition 1 (Prediction Market Equilibrium without Social Networks) *In a non-networked prediction market, the equilibrium prediction market price is given by*

$$P^* = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[V|I_i] = \frac{\rho_V}{\rho_\varepsilon+\rho_V} V_0 + \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V} S_i$$

$$= \frac{\rho_V}{\rho_\varepsilon+\rho_V} V_0 + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V} V + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V} \overline{\varepsilon},$$

where $\overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$*, and the equilibrium position for each participant is* $x_i^* = \frac{\rho_\varepsilon(\varepsilon_i - \overline{\varepsilon})}{2\gamma}$.

The market price P^* is the forecast/estimator generated from the prediction market. Proposition 1 shows the information aggregation mechanism of a prediction market where P^* reflects participants' diverse expectations, $\frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[V|I_i]$. According to the equation $P^* = \frac{\rho_V}{\rho_\varepsilon+\rho_V} V_0 + \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V} S_i$, the information-aggregation mechanism places weights on the public information and each individual's private information. The weight on the public information in a non-networked prediction market is given by:

$$W_{NP} = \frac{\rho_V}{\rho_\varepsilon+\rho_V} \Big/ \left(\frac{\rho_V}{\rho_\varepsilon+\rho_V} + \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}\right) = \frac{\rho_V}{\rho_\varepsilon+\rho_V}. \quad (5)$$

A natural question arises: what is the socially efficient weight on the public information? Suppose that we have an ideal scenario. The corporate manager can perfectly assess all prediction market participants' private information without relying on a corporate prediction market. Then, the corporate manager's best guess is her conditional expectation of V, which is a weighted average of the public information and all private signals:

$$P_m = \mathbf{E}[V|I_m] = \frac{\rho_V}{n\rho_\varepsilon+\rho_V} V_0 + \sum_{i=1}^{n} \frac{\rho_\varepsilon}{n\rho_\varepsilon+\rho_V} S_i,$$

where I_m is the managers' information set. Therefore, the efficient weight on the public information (first best) is:

$$W_m = \frac{\rho_V}{n\rho_\varepsilon+\rho_V} \Big/ \left(\frac{\rho_V}{n\rho_\varepsilon+\rho_V} + \sum_{i=1}^{n} \frac{\rho_\varepsilon}{n\rho_\varepsilon+\rho_V}\right) = \frac{\rho_V}{n\rho_\varepsilon+\rho_V}. \quad (6)$$

Comparing equation 6 with equation 5, we find that $W_{NP} \geq W_m$, which implies that the weight on the public information in a non-networked prediction market is larger than the efficient weight. This shows that the problem of overweighting the public information still exists in a corporate prediction market.

Following the prior literature (Lamberson and Page 2012; Davis-Stober et al. 2015), we use the mean squared error (MSE) to measure the prediction market performance or prediction accuracy. The MSE of an estimator measures the average of the squares of the "errors." The larger the MSE is, the less accurate the forecast generated from the prediction market is. In a non-networked prediction market, the MSE of the forecast P^* is given by:

$$\text{MSE}(P^*) = \mathbf{E}[(V - P^*)^2] = \frac{\rho_V}{(\rho_\varepsilon+\rho_V)^2} + \frac{\rho_\varepsilon}{n(\rho_\varepsilon+\rho_V)^2}, \tag{7}$$

and the MSE in the ideal scenario in which the manager can assess all prediction market participants' private information is:

$$\text{MSE}(P_m) = \mathbf{E}[(V - P_m)^2] = \frac{1}{n\rho_\varepsilon+\rho_V} \leq \text{MSE}(P^*). \tag{8}$$

From equation 7, we have the following proposition:

Proposition 2 (Comparative Statics on MSE) *In a non-networked prediction market, the MSE of the forecast P^* decreases with the number of prediction market participants, n, and the precision of private signals, ρ_ε. If $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{n-2}{n}$, the MSE increases with the precision of the common prior (public information); if $\frac{\rho_V}{\rho_\varepsilon} > \frac{n-2}{n}$, the MSE decreases with the precision of public information.*

The implications of this proposition are as follows. First, the result of comparative statics of n on MSE is straightforward and consistent with our intuition. The prediction market accuracy increases with the number of participants. This result is reminiscent of the power of the wisdom of crowds: "Under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them" (Surowiecki 2004, page 41). The prediction errors are cancelled out when the number of participants is large. If $n \to \infty$, the MSE converges to $\rho_V/(\rho_\varepsilon + \rho_V)^2$.

More importantly, we find that in a non-networked prediction market, increased precision of private information always enhances the prediction market accuracy. However, the impact of the public information precision is intriguing: When the precision of public information, ρ_V, is relatively large to the precision of private information, ρ_ε, greater precision of the public information increases prediction market accuracy. However, when ρ_V is relatively small to ρ_ε, greater precision of public information is detrimental to prediction market accuracy.

In a prediction market with a very large number of participants ($n \to \infty$), our result can be simplified as follows: If $\rho_V \geq \rho_\varepsilon$, the MSE decreases with ρ_V; if $\rho_V < \rho_\varepsilon$, the MSE increases with ρ_V. This result is surprising in the sense that when we consider each participant's decision making problem, more precise information is generally beneficial to the participant no matter whether the information is private (available only to participant i) or public (shared by all participants). However, it is not always the case that greater precision of the public information is desirable in terms of prediction market performance.

The key insight from our model is that increased precision of public information is beneficial only when it is precise. The underlying intuition is in line with the overweighting effect documented in the extant literature (Morris and Shin 2002; Angeletos and Pavan 2007). In our prediction market, the public information conveys useful information on the uncertain event, V. On the other hand, everyone receives the same public information. The detrimental impact arises from the fact that the information-aggregation mechanism places a larger-than-efficient weight on the public information: $W_{NP} \geq W_m$. More specifically, when each participant forms her expectation of the uncertain event, $\mathbf{E}[V|I_i]$, she will give certain weight to the public information as her best guess. Then, the prediction market aggregates all participants' expectations using a security-trading mechanism. Since every participant gives certain weight to the public information, the prediction market forecast will overreact to the public information because the public information is counted multiple times. Any noise contained in the public information will be magnified by overweighting the public information. Therefore, when the public information is less precise, we are more likely to observe that greater precision of public information lowers the prediction market accuracy. Note that the impact of increased precision of public information in a non-networked prediction market is given by:

$$\frac{\partial}{\partial \rho_V}\mathrm{MSE}(P^*) = \underbrace{\frac{(n-2)\rho_\varepsilon}{n(\rho_\varepsilon+\rho_V)^3}}_{\substack{\text{Detrimental} \\ \text{Effect}}} - \underbrace{\frac{\rho_V}{(\rho_\varepsilon+\rho_V)^3}}_{\substack{\text{Beneficial} \\ \text{Effect}}}, \tag{9}$$

where the first term indicates the detrimental effect of public information on the prediction market performance due to the overweighting problem, and the second term indicates the beneficial effect of public information since it conveys useful information about V. The impact of increased precision of the public information depends on the relative strength of these two effects.

Our following numerical example further illustrates the implications of Proposition 2 and gives a visualization of the regions of ρ_V and ρ_ε, for which the prediction market accuracy measured by the MSE is increasing or decreasing in ρ_V. In this numerical example, we set the number of prediction market participants, $n = 50$. The results are robust when we vary n. Figure 1(a) displays the MSE for different values of the precisions of private information and public information. Figure 1(b) depicts the contour lines of the MSE. The whole region in Figure 1(b) can be divided by the marginal line $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n} = \frac{48}{50}$. In Region I, greater precision of the public information increases the prediction market accuracy. However, in Region II, greater precision of the public information is detrimental to prediction market accuracy.

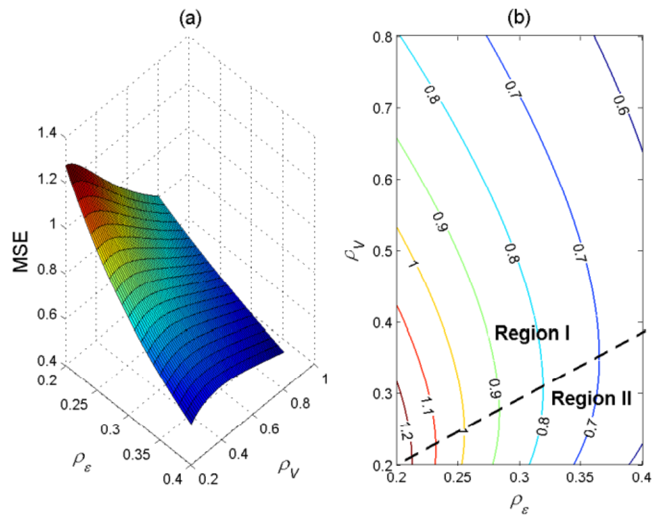


Figure 1. The Impact of Public Information and Private Information on Prediction Market Performance (Non-Networked Case), $n = 50$.

Figure 2 shows how the number of participants affects the size of Region II. The dotted line is $\frac{\rho_V}{\rho_\varepsilon} = \frac{1}{3}$, which corresponds to $n = 3$. Similarly, the solid and dashed lines represent $\frac{\rho_V}{\rho_\varepsilon} = \frac{8}{10}$ and $\frac{\rho_V}{\rho_\varepsilon} = \frac{48}{50}$, respectively. As the number of participants increases, the marginal line will move up and converge to $\frac{\rho_V}{\rho_\varepsilon} = 1$. In other words, when the number of participants is larger, increased precision of public information is more likely to be detrimental—that is, Region II is larger, and Region I is smaller. If $n = 3$, the condition for a detrimental effect of the public information is $\rho_\varepsilon \geq 3\rho_V$. If $n = 50$, the condition for a detrimental effect of the public information is $\rho_\varepsilon \geq \frac{25}{24}\rho_V$. More generally, we have the following proposition:

Proposition 3 *In a non-networked prediction market, increased precision of public information is more likely to be detrimental to the prediction market performance as* n *increases.*

The intuition of Propostion 3 can be derived by examining equation 9. The beneficial effect in equation 9 does not depend on n, but the detrimental effect increases with n. Actually, the overweighting problem becomes more serious as n increases. From equations 5 and 6, we find that the weight difference, $W_{NP} - W_m$, increases with n. Therefore, increased precision of public information is more likely to be detrimental to the prediction market performance as n increases. Note that in the practice of corporate prediction markets, the number of participants typically varies from 20 to 50 or even larger (Chen and Plott 2002). When $n = 20$, the condition for a detrimental effect of the public information is $\rho_\varepsilon \geq \frac{10}{9}\rho_V$. It means that whenever the precision of the private signal is greater than $10/9$ of the precision of the public information, prediction market accuracy is decreasing in ρ_V. This condition is likely to hold in reality because corporate prediction market participants are internal employees, and they may have more precise insider information (private signals) than the public information (Qiu, Rui, and Whinston 2014b).

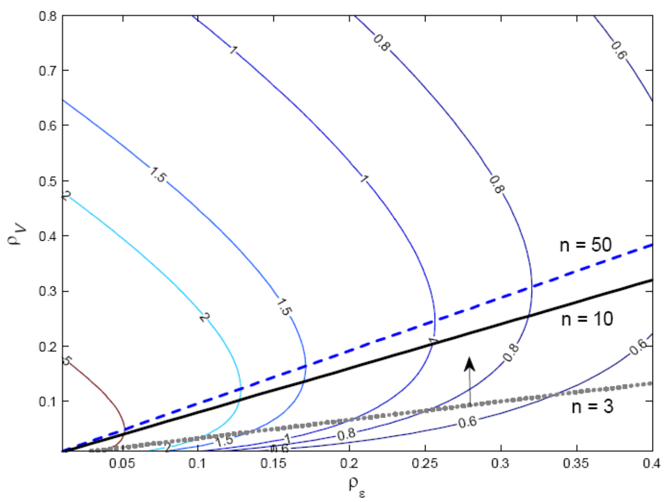


Figure 2. The Impact of the Number of Prediction Market Participants

In summary, in the corporate prediction market design, the number of participants is critical not because it can directly affect the prediction market performance. According to equation 7, the marginal beneficial effect of prediction market size decreases as n increases. Prior empirical studies have also confirmed that the marginal beneficial effect of prediction market size on prediction market accuracy is small when the number of participants exceeds 20 (McHugh and Jackson 2012). The real reason why we should care about n is its impact on the condition for a detrimental effect of the public information: $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{n-2}{n}$. Proposition 2 can inform managers the market conditions under which increased precision of public information is not beneficial. In the design of a corporate prediction market, managers should exercise caution in how much and how precise the public information they reveal. When the public information is relatively noisy, revealing more precise public information to prediction market participants may hurt prediction market accuracy.

3.3 A Mixture Model of REE and DO

We develop a mixture model that nests both the REE and DO approaches where, among total *n* participants, *m* participants use the DO approach to make inferences, and *n* – *m* participants use the REE approach to make inferences. We denote the set of DO traders as C_{DO}, which contains *m* partcipants. The rest *n* – *m* participants are REE traders. A DO trader $i \in C_{DO}$ makes a Bayesian inference using her private signal and the common prior, as described in Section 3.2.

If participant i is an REE trader, then her information set I_i is her private signal S_i as well as the price function $P^*(V)$. We denote the set of REE traders as C_{REE}, which contains *n* - *m* partcipants. The central tenet of the REE literature (Grossman and Stiglitz 1980; Kyle 1985; O'hara 1995) is that the market price is a function of the fundamental value V. Hence, a fully rational consumer is able to learn from the price function $P^*(V)$. Intuitively speaking, $P^*(V)$ is self-fulfilling: When participants think prices as being generated by $P^*(V)$, they will act in such a way that the market clears at $P^*(V)$. Mathemetically speaking, it is essentially a fixed-point problem. Following the finance and economics literature (Grossman and Stiglitz 1980; Kyle 1985; O'hara 1995), we solve the fixed-point problem by assuming that an REE particpant forms a linear conjecture on the equlibrium price function:

$$P^*(V) = a + bV + c\bar{\varepsilon},$$

where $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$, and *a*, *b*, and *c* are three constants to be dertermined. Recall that for both types of traders, the optimal trading position is given by equation 4. Using the market clearing condition, $\sum_{i \in C_{DO}} x_i^* + \sum_{j \in C_{REE}} x_j^* = 0$, we obtain the following proposition:

Proposition 4 (Prediction Market Equilibrium with both DO and REE Traders) *In a prediction market with both DO and REE traders, the equilibrium prediction market price is given by*

$$P^* = a + bV + c\bar{\varepsilon},$$

where $a = \frac{\rho_V}{(n+1-m)\rho_\varepsilon + \rho_V} V_0$, *and* $b = c = \frac{(n+1-m)\rho_\varepsilon}{(n+1-m)\rho_\varepsilon + \rho_V}$.

From Proposition 4, we find that the weight on public information in a market with both DO and REE traders is $\frac{\rho_V}{(n+1-m)\rho_\varepsilon + \rho_V}$. From equation 6, the socially efficent weight is $\frac{\rho_V}{n\rho_\varepsilon + \rho_V}$. When there is more than one DO trader (i.e., $m > 1$), the problem of overweighting the public information still exists in a prediction market with both DO and REE traders, and the overweighting problem is most serious when all prediction market participants are DO traders, $m = n$. We also compute the MSE of P^* in a prediction market with both DO and REE traders:

$$\text{MSE}(P^*) = \mathbf{E}[(V - P^*)^2] = \frac{\rho_V}{[(n+1-m)\rho_\varepsilon + \rho_V]^2} + \frac{\rho_\varepsilon (n+1-m)^2}{n[(n+1-m)\rho_\varepsilon + \rho_V]^2}.$$

Note that when $m = n$, the MSE above will convert to the MSE in a prediction market where all participants are DO traders. Based on the MSE, we obtain the following two proposistions.

Proposition 5 *In a prediction market with both DO and REE traders, when* $m \geq 1$, *the MSE of the forecast* P^* *increases with the number of DO traders, m.*

This propostion is consistent with our explanations on the overweight problem of public information. As the number of DO traders increases, the problem of overweighting public information will become more serious, and hence the prediction market performance will decrease (the MSE will increase).

Proposition 6 (Comparative Statics on MSE) *In a prediction market with both DO and REE traders, the MSE of the forecast* P^* *decreases with the number of prediction market participants,* n, *and the precision of private signals,* ρ_ε. *When* $m \leq \frac{n+2}{2}$, *the MSE of the forecast* P^* *always decreases with the precision of public information. When* $m > \frac{n+2}{2}$, *if* $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{(2m-2-n)(n+1-m)}{n}$, *the MSE increases with the precision of public information; and if* $\frac{\rho_V}{\rho_\varepsilon} > \frac{(2m-2-n)(n+1-m)}{n}$, *the MSE decreases with the precision of* *public information.*

The results in a prediction market that consists of both DO and REE traders are similar to those in a prediction market that consists of only DO traders. Increased number of participants and increased precision of private information always enhance prediction market accuracy; however, the impact of public information precision is conditional on the number of DO traders, m. If $m > \frac{n+2}{2}$, the impact of the public information precision depends on the precisions of private and public information.

As we have shown in Proposistions 4, 5, and 6, the key analytical results in the model that nests both the REE and DO approaches are qualitatively similar to those in a pure DO model: (i) the public information is overweighted, and (ii) increased precision of public information is not always beneficial to prediction market perforamance. In the reminder of the paper, analyzing a socially embedded prediction market, we will focus on the pure DO model because the model nesting both the REE and DO approaches complicates our analyses and does not add additional analytical insights.

4 A Socially Embedded Prediction Market

In this section, we examine the impact of information exchange in social networks on the prediction market performance in a pure DO framework. In our benchmark model, participants are isolated in the sense that they receive conditionally independent private information and cannot communicate with each other. However, in real corporate prediction markets, participants may receive information from each other in different forms (Qiu, Rui, and Whinston 2014a). For instance, they may chat about their prediction tasks during their coffee breaks. Cowgill, Wolfers, and Zitzewitz (2009) found correlated tradings among employees who sit within a few feet of one another and employees with social or work relationships in Google's prediction markets, which suggest that prediction market participants may share private information with their social connections. More specifically, in our socially embedded prediction markets, each participant receives a private signal and exchanges information with their friends in a social network. The social network $\Gamma = (N, L)$ is given by a finite set of nodes $N =$

$\{1,2,\ldots,n\}$ and a set of links $L \subseteq N \times N$. Each node represents a participant in the prediction market. The social connections between the participants are described by an $n \times n$ dimensional matrix denoted by $g \in \{0,1\}^{n\times n}$, such that:

$$g_{ij} = \begin{cases} 1, & if \quad (i,j) \in L \\ 0, & otherwise \end{cases},$$

where $g_{ij} = 1$ implies that participants i and j are friends; otherwise, they are not. Let $N_i(g) = \{j \in N: g_{ij} = 1\}$ represent the set of friends of participant i. The degree of participant i is the number of participant i's friends: $k_i(g) = \#N_i(g)$. Following the prior literature (Ozsoylev and Walden 2011; Han and Yang 2013), we assume that the social network is undirected and that prediction market participants can freely communicate their private signals to others that are connected to them in the network.[15] In other words, participant i can observe her friends' signals, S_j, $j \in N_i(g)$, and take them into account in her inference process.

In the DO framework, each participant completely igores others' information contained in market prices. However, in our social network setup, we assume that DO participants consider the signals from their friends in the inference process. Actually, these two assumptions are compatible. As we pointed out, DO can be motivated by behavioral biases, such as bounded rationality and limited computational capacity, or heterogeneous priors. In our study, we adopt the first explanation: A prediction market participant lacks the *ability* to extract other participants' information from market prices.[16] However, if other participants' signals are directly given to her, she should have no problems using the signals. The rationality requirement for using available signals from friends is much lower than extracting other participants' information from market prices, because drawing inferences from market prices requires complete knowledge of the market clearing process and correct conjectures on equlibrium prices. In our

---

[15] For simplicity, we assume the network is undirected, but the results also hold for directed networks.

[16] As Hong and Stein (2003, p. 491) pointed out, "the differences of opinion can be thought of as reflecting a type of bounded rationality in which investors are simply unable to make inferences from prices."

bounded rationality framework, the fact that a participant ignores others' information contained in market prices is not because she always wants to ignore others' information. The underlying deep reason is the constraint on participants' information-processing abilities: They are unable to extract others' information from market prices because they have limited cognitive abilities to process information (Kahneman 2003).

The key difference between the information contained in market prices and signals passed from friends is how information is presented and displayed. In an analytical model, Hirshleifer and Teoh (2003) assumed that financial information that is presented in a salient, easily processed form can be absorbed more easily by traders than information that is less salient and difficult to process because traders have limited attention and processing power. In our context, other participants' information contained in market prices is less salient and difficult to process.[17] Peng (2005) pointed out that learning from prices is not free, since doing so requires knowledge of the structure of the market. Traders have limited time and attention to process information, and the capacity constraint limits the amount of information that she can process. An important argument in Hirshleifer and Teoh (2003) is that limited information processing capacity tends to induce participants to use information that is presented in salient, easily processed form (in our context, it refers to the signals from friends) rather than non-salient or hard-to-process information (in our context, it refers to the information contained in market prices).[18]

Additionally, experimental evidence in prior literature suggests that extracting information from

[17] Hong and Stein (1999) proposed a model based on investors with bounded rationality to explain the momentum phenomenon in financial markets. They placed constraints on traders' information-processing abilities and assumed a group of boundedly rational agents, "newswatchers": Each newswatcher observes her private information, but fails to extract other newswatchers' information from prices. More specifically, the newswatchers make forecasts based on signals that they privately observe about future fundamentals; their limitation is that they do not condition on prices. Hong and Stein (1999) argued that this assumption can be motivated by bounded rationality, a plausible and intuitively appealing explanation: Traders are unable to use market prices to form more sophisticated forecasts.

[18] It is worth noting that the rationality requirement of extracting useful information from market prices in an REE is higher than that of drawing inferences from a given price. In an REE, each trader needs to correctly conjecture the equilibrium price (which will be determined by each trader's trading position), and then draws inferences from the price and makes her trading decision. Therefore, the rationality requirement of extracting useful information from market prices in an REE consists of two critical steps: (i) forming a correct expectation on the equilibrium price, and (ii) drawing inferences from the given price. Apparently, forming a correct expectation on the equilibrium price is challenging work. In an analytical model, Vives (1993) showed that the speed of learning to form correct expectations (the speed of converging to correct rational expectations) is very slow. In contrast, when each participant considers their friends' signals in our social-network-embedded prediction markets, the friends' signals are given to each participant directly. The rationality requirement of considering others' signals is much lower than that of extracting useful information from market prices because each participant is not required to form correct expectations on market prices.

prices is more difficult than using available information directly. The time and attention needed to process financial information contained in prices is non-trivial. The empirical findings from laboratory markets show that traders rarely extract information that is available in prices. For instance, Bloomfield, Tayler, and Zhou (2009) tested a key assumption in Hong and Stein (1999) in a laboratory experiment—traders' inability to draw inferences from the market price—and found evidence supporting that the traders fail to infer other traders' information from market prices. In another laboratory experiment, Corgnet, DeSantis, and Porter (2015) found that their data can be best explained by the model in which traders do not infer other traders' information from market prices but apply Bayes' rule to compute the expected value of the asset given their own information.

To make our model analytically tractable, we consider three special cases of a general social graph g: (i) a regular social network without homophily, where every participant has the same degree k (Jackson 2008). In this case, we assume the private signal errors to be independent across all participants; (ii) a regular social network with homophily. Homophily is a typical phenomenon observed in social networks in that there are inherent similarities in friends' personal characteristics (Aral and Walker 2011; Gu et al. 2014; Bapna and Umyarov 2015). In this case, we assume that the errors of private signals are positively correlated; and (iii) a heterogenous social network where a participant has either degree 0 or degree k (the analytical results of this case can be found in online appendix B). Although oversimplified, the network structure in these three cases reflects some fundamental features of typical social networks in reality and enables us to draw analytical insights. For instance, Cases (i) and (ii) are more balanced social networks without degree heterogeneity and reflect a flat organization since no one is located in the center of the network. This is similar to the assumption in Ozsoylev and Walden (2011): No agent is informationally superior and possesses too much information. Case (iii) is a more heterogenous social network and may reflect a socially embedded prediction market consisting of both well-connected employees who have high social skills and isolated employees who have low social skills. Because the case of a general social graph is not analytically tractable, we run numerical simulations and show that our analytical insights remain robust when the underlying networks are more complicated (e.g., the Erdos–Renyi random graph, the Gilbert graph, the "small world" graph, and the preferential attachment graph) in Section 4.3.

4.1 A Regular Social Network without Homophily

In a regular network, each participant has k friends and can receive private signals of her friends. Therefore, a participant i's information set, I_i, includes her private signal, her friends' private signals (k signals), and the common prior. She makes an inference as follows:

$$\mathbf{E}[V|I_i] = \frac{\rho_V}{(k+1)\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}S_i + \sum_{j\in N_i(g)} \frac{\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}S_j,$$

$$\mathbf{Var}[V|I_i] = 1/[(k+1)\rho_\varepsilon + \rho_V].$$

Similarly, participant i's position is given by equation 4, and the equilibrium prediction market price P^* is determined by the market clearing condition, $\sum_{i=1}^{n} x_i^* = 0$.

The following proposition characterizes the equilibrium of a prediction market with a regular social network.

Proposition 7 (Prediction Market Equilibrium in a Regular Social Network) *In a prediction market with a regular social network, the equilibrium prediction market price is given by*

$$P^* = \frac{\rho_V}{(k+1)\rho_\varepsilon+\rho_V}V_0 + \frac{(k+1)\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}V + \frac{(k+1)\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}\overline{\varepsilon},$$

where $\overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$*, and the equilibrium position for each participant is*

$$x_i^* = \frac{\rho_\varepsilon}{2\gamma}\left[\varepsilon_i + \sum_{j\in N_i(g)} \varepsilon_j - (k+1)\overline{\varepsilon}\right].$$

Proposition 7 indicates that in a prediction market with a regular social network, the weight on public information is given by:

$$W_{RS} = \frac{\rho_V}{(k+1)\rho_\varepsilon+\rho_V} \le W_{NP}. \tag{10}$$

If we compare equation 10 with equations 5 and 6, we find that $W_m \le W_{RS} \le W_{NP}$, where $W_m = W_{RS}$

when $k = n - 1$, and $W_{RS} = W_{NP}$ when $k = 0$. The implication is that social interaction among prediction market participants can correct the problem of overweighting the public information. When the level of social interactions reaches the maximum (a complete or a fully connected social network, $k = n - 1$), the weight on the public information is efficient in a prediction market with a regular network.

Then, we compute the MSE of P^* in a prediction market with a regular social network:

$$\text{MSE}(P^*) = \mathbf{E}[(V - P^*)^2] = \frac{\rho_V}{[(k+1)\rho_\varepsilon + \rho_V]^2} + \frac{\rho_\varepsilon (k+1)^2}{n[(k+1)\rho_\varepsilon + \rho_V]^2}. \tag{11}$$

Note that when $k = 0$, the MSE in a prediction market with a regular social network given by equation 11 will convert to equation 7, the MSE in a non-networked prediction market. In general, we can consider a non-networked prediction market as a special case of a prediction market with a regular network ($k = 0$).

Comparing the MSE in a prediction market with a regular social network with that in a non-networked prediction market, we obtain the following proposition:

Proposition 8 (MSE Comparison: No Network vs. Regular Network) *The MSE in a non-networked prediction market is greater than the MSE in a prediction market with a regular social network.*

Proposition 8 shows that a prediction market with a regular social network outperforms a prediction market without social networks. As we have explained before, the problem of a non-networked prediction market is that the information-aggregation process will count the public information multiple times, and therefore magnify the noise contained in the public information. The existence of a regular social network facilitates private information exchange among participants, which effectively puts a larger weight on the private information. In a regular network, each participant will receive her friends' private signals, and her own private signal will be received by k friends. In other words, each private signal will be counted k times when participants' predictions are aggregated in the prediction market. Such multiple counting of private information is beneficial to the prediction market performance because it can correct the bias toward the public information caused by overweighting the public information. Essentially, the advantage of embedding a social network is to use the *multiple counting of private information* to neutralize the harmful effect from the *multiple counting of public information*.

We examine the effect of the level of social interactions, k, on the prediction market performance in the following proposition:

Proposition 9 (Impact of Social Interaction Level) *The MSE in a prediction market with a regular network decreases with* k.

Proposition 9 shows that the prediction market performance increases with the level of social interactions, k. In practice, a manager may want to encourage social interactions among participants to improve prediction market accuracy. The intuition is similar to that in Proposition 5. As the level of social interactions, k, increases, the bias toward the public information will be corrected to a larger extent, and the weight on the public information in the information-aggregation process will be closer to the efficient value.

To examine the impact of the precision of public and private information, we have the following proposition:

Proposition 10 (Comparative Statics on MSE) *In a prediction market with a regular social network, the MSE of the forecast* P^* *decreases with the number of prediction market participants,* n, *and the precision of private signals,* ρ_ε. *If* $\frac{\rho_V}{\rho_\varepsilon} \leq (k+1)\left[\frac{n-2(k+1)}{n}\right]$, *the MSE increases with the precision of public information; if* $\frac{\rho_V}{\rho_\varepsilon} > (k+1)\left[\frac{n-2(k+1)}{n}\right]$, *the MSE decreases with the precision of public information.*

The results in a prediction market with a regular social network are similar to those in a non-networked prediction market. Increased precision of private information always enhances prediction market accuracy, but the impact of public information precision depends on the relative precision of private information versus public information. When ρ_ε is relatively small to ρ_V, greater precision of the public information increases the prediction market accuracy. When ρ_ε is relatively large to ρ_V, greater precision of the public information decreases the prediction market accuracy. In a socially embedded prediction market, the prediction performance depends on not only the precisions of private and public information but also the level of social interactions, k.

An interesting observation from Propositions 8, 9, and 10 is that a socially embedded prediction market with low precision of the private information may perform as well as a non-networked prediction market with high precision of the private information. A managerial implication of this result is about the selection of prediction market participants. In general, an internal employee has two types of skills: "work skills" and "social skills." In our context, the level of work skills refers to the ability to acquire precise private information (knowledge creation and information production) and is measured by ρ_ε. In contrast, the level of social skills refers to the ability to communicate and share information with colleagues (knowledge transfer and information communication) and is measured by k. Intuitively, a manager should select employees who have a high level of work skills (ρ_ε) as prediction market participants. This is also consistent with Proposition 10. However, Proposition 9 shows that the level of social skills (k) also matters when we consider the prediction market performance. A group of participants who have a medium level of work skills but a high level of social skills may outperform those who have a high level of work skills but a low level of social skills. We provide a numerical example in online Appendix C.

Regarding the precision of public information, in a prediction market with a regular network, the impact of increased precision of public information is given by:

$$\frac{\partial}{\partial \rho_V}\mathrm{MSE}(P^*) = \underbrace{\frac{[n-2(k+1)](k+1)\rho_\varepsilon}{n[(k+1)\rho_\varepsilon+\rho_V]^3}}_{\substack{\text{Detrimental}\\ \text{Effect}}} - \underbrace{\frac{\rho_V}{[(k+1)\rho_\varepsilon+\rho_V]^3}}_{\substack{\text{Beneficial}\\ \text{Effect}}}. \tag{12}$$

As we have stated, the impact of increased precision of public information depends on the relative strength of the detrimental and beneficial effects. The beneficial effect in equation 12 does not depend on n, but the detrimental effect increases with n. Therefore, we should expect that increased precision of public information is more likely to be detrimental to the prediction market performance as n increases. As for the level of social interactions, k, we have the following proposition:

Proposition 11 *In a prediction market with a regular network, increased precision of public information is more likely to be detrimental to the prediction market performance as* n *increases. When* $n \geq 4(k+1)$, *increased precision of public information is more likely to be detrimental to the prediction market performance as* k *increases; when* $n < 4(k+1)$, *increased precision of public information is less likely to be detrimental to the prediction market performance as* k *increases. Specifically, if* $n < 2(k+1)$, *the prediction market performance will always increase with the precision of public information.*

As in the previous section, we define Region I as the range of market conditions in which increased precision of public information enhances the prediction market accuracy and Region II as the range of market conditions in which increased precision of public information decreases the prediction market accuracy. Proposition 11 indicates the conditions in which Region II becomes larger as the level of social interactions, k, increases in a prediction market with a regular social network. When n is large relative to k, the size of Region II increases with k. When n is small relative to k, the size of Region II decreases with k.

To provide some additional intuition, we conduct a numerical analysis and visualize Proposition 11. In Figure 3, we set $n = 50$, and vary degree k. The solid line represents $\frac{\rho_V}{\rho_\varepsilon} = (k+1)\left[\frac{n-2(k+1)}{n}\right]$, where $k = 1, 10, 20$, and 25, and the dashed line represents the marginal line in the case of no social networks ($k = 0$): $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n}$. According to Proposition 11, if $n \geq 2(k+2)$, Region II should be larger in a prediction market with a regular social network than in a non-networked prediction market. In Figure 3(a), the marginal line moves up as k increases from 0 to 1, which suggests that Region II is larger when a prediction market is embedded in a regular social network. In Figures 3(a) and (b), we find that Region II becomes larger as k increases from 1 to 10. However, the size of Region II shrinks as k increases from 10 to 20 in Figures 3(b) and (c). Eventually, when $k = 25$, Region II does not exist (only Region I left) in Figure 3(d). These results are consistent with Proposition 11: (i) When $n \geq 4(k+1)$, increased precision of public information is more likely to be detrimental to the prediction market performance as k increases; (ii) when $n < 4(k+1)$, increased precision of public information is less likely to be detrimental to the prediction market performance as k increases; and (iii) when $n < 2(k+1)$, the prediction market performance will always increase with the precision of public information.

Proposition 11 has important managerial implications to corporate prediction market designers. First, although a socially embedded prediction market can improve the prediction market performance, corporate prediction market designers should pay more attention to the public information disclosure in a socially embedded prediction market. Even if a manager has obtained some additional information that can increase the precision of public information, it may not be a good idea to disclose it to all prediction market participants. In a socially embedded prediction market, the size of Region II could be much larger than in a non-networked prediction market under some circumstances. For instance, in Figure 3(b), where $k = 10$ and $n = 50$, increased precision of public information is detrimental to the prediction market performance if $\frac{\rho_V}{\rho_\varepsilon} \leq 6.16$ (the solid line). However, in a non-networked environment, increased precision of public information is detrimental to the prediction market performance if $\frac{\rho_V}{\rho_\varepsilon} \leq 0.96$ (the dashed line). Additionally, when $n \to \infty$ (a large prediction market with many participants), increased precision of public information is detrimental to prediction market performance if $\frac{\rho_V}{\rho_\varepsilon} \leq k+1$ in a regular network case, whereas the condition is $\frac{\rho_V}{\rho_\varepsilon} \leq 1$ in a non-networked case. For reasonable parameter choices in reality, the detrimental effect of public information is much more likely to occur in a socially embedded prediction market than in a non-networked case.

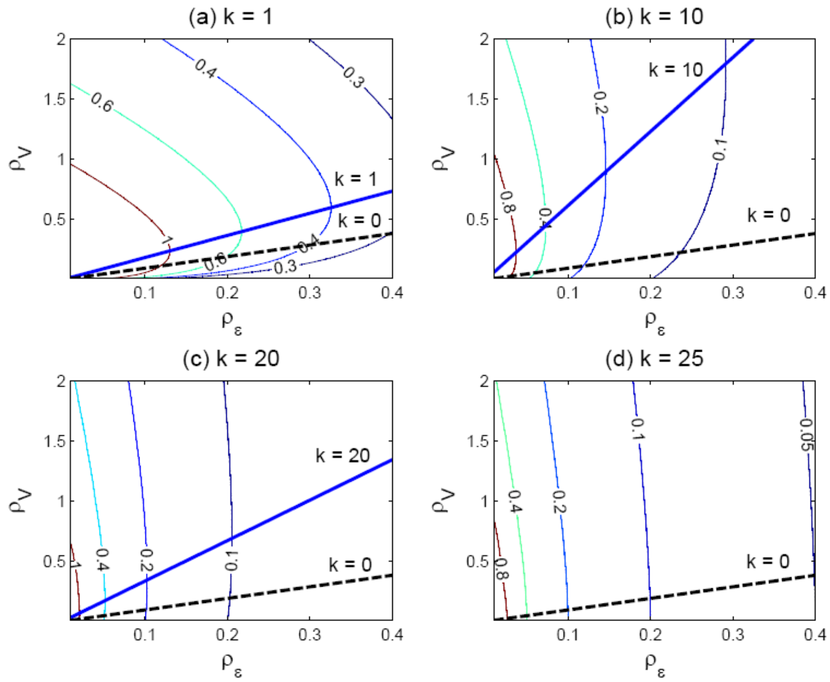


Figure 3. The Impact of Public Information and Private Information on Prediction Market Performance (Regular Network), $n = 50$.

4.2 A Regular Social Network with Homophily

In the previous analysis, we assume the private signal errors to be independent across all participants. In our context, participants connected through a social network may be similar to each other, and therefore may have similar information sources when formulating their private information, such as reading the same reports and news articles, working in the same department of the company, having similar educational and working experiences and so on.[19] In this section, we relax the assumption of independent private information and assume that the errors of friends' private signals are positively correlated to reflect the plausible existence of homophily.

For analytical tractability, we first look at a regular social network with $k_i(g) = k = 1$. It means

[19] If prediction market participants are selected from different departments of a company, they are likely to have diverse information sources, and hence their private signals are independent. If prediction market participants are selected from the same department, their information could be highly correlated because they are exposed to similar information sources. For instance, in General Electric (GE)'s prediction market, participation included employees representing 150 business segments from 42 countries. Private signals of employees from different countries and different business segments are less likely to be correlated. See http://www.consensuspoint.com/wp-content/themes/radius/whitepapers/GE_Casestudy.pdf. As argued in Keuschnigg and Ganser (2016), crowd wisdom does not only depend on the prediction ability/precision of agents, but also on the information diversity. We conduct a simulation analysis on the trade-off between information diversity and information precision, and the results can be found in online appendix F.

that each participant has one friend and can receive the private signal of her friend. We will examine the impact of homophily in more complicated social networks using simulations in Section 4.3. Because of the homphily effect, the error of partcipant *i*'s private signal, ε_i, is correalted with her friend *j*'s priavate signal error ε_j:

$$\begin{pmatrix} \varepsilon_i \\ \varepsilon_j \end{pmatrix} \sim N(0, \Sigma), \Sigma = \begin{pmatrix} 1/\rho_\varepsilon & \delta/\rho_\varepsilon \\ \delta/\rho_\varepsilon & 1/\rho_\varepsilon \end{pmatrix},$$

where Σ is the covariance matrix for participant i's and participant j's signal errors, and δ is the correlation coefficient. In order to capture the homophily effect, we assume that $0 \le \delta \le 1$. Participant *i* makes an inference as follows:

$$\mathbf{E}[V|I_i] = \frac{\rho_V}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V} V_0 + \frac{\rho_\varepsilon}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V} \frac{1}{1+\delta} S_i + \frac{\rho_\varepsilon}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V} \frac{1}{1+\delta} S_j, \ \mathbf{Var}[V|I_i] = 1/\left[\frac{2}{1+\delta}\rho_\varepsilon + \rho_V\right].$$

The following proposition characterizes the equilibrium of a prediction market with a homophily social network.

Proposition 12 (Prediction Market Equilibrium under Homophily) *In a prediction market under homophily, the equilibrium prediction market price is given by*

$$P^* = \frac{\rho_V}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V} V_0 + \frac{2\rho_\varepsilon}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V} \frac{1}{1+\delta} V + \frac{2\rho_\varepsilon}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V} \frac{1}{1+\delta} \bar{\varepsilon}.$$

Proposition 12 shows that in a prediction market under homophily, the weight on public information is $W_H = \frac{\rho_V}{\frac{2}{1+\delta}\rho_\varepsilon + \rho_V}$. According to our previous analysis, in a prediction market with a regular social network (no homophily), the weight on public information when $k = 1$ is $W_{RS} = \frac{\rho_V}{(k+1)\rho_\varepsilon + \rho_V} = \frac{\rho_V}{2\rho_\varepsilon + \rho_V} \le W_H$. The implication is that even though social interaction among prediction market participants can correct the problem of overweighting the public information, the homophily effect tends to weaken the correction because the friend's information is less useful under homophily. In an extreme case with perfect correlation ($\delta = 1$), $W_H = W_{NP}$, which means that social interactions cannot correct the overweighting problem at all when the friend's private signal does not contain additional value.

Then, we compute the MSE of P^* in a prediction market under homophily:

$$\text{MSE}(P^*) = \mathbf{E}[(V - P^*)^2] = \frac{\rho_V}{\left[\frac{2}{1+\delta}\rho_\varepsilon + \rho_V\right]^2} + \frac{\rho_\varepsilon \frac{4}{1+\delta}}{n\left[\frac{2}{1+\delta}\rho_\varepsilon + \rho_V\right]^2},$$

and we obtain the following propositions:

Proposition 13 (Impact of Homophily) *The MSE in a prediction market under homophily increases with* δ*.*

Proposition 14 (Comparative Statics on MSE) *In a prediction market under homophily, the MSE of the forecast* P^* *decreases with the number of prediction market participants,* n*, and the precision of private signals,* ρ_ε*. If* $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{2}{1+\delta}\left[\frac{n-4}{n}\right]$*, the MSE increases with the precision of public information; if* $\frac{\rho_V}{\rho_\varepsilon} > \frac{2}{1+\delta}\left[\frac{n-4}{n}\right]$*, the MSE decreases with the precision of public information.*

Proposition 13 suggests that the homophily effect tends to be detrimental to the prediction market performance. Proposition 14 further shows that the impact of increased precision of public information depends on the information correlation coefficent δ, which measures the effect of homophily.

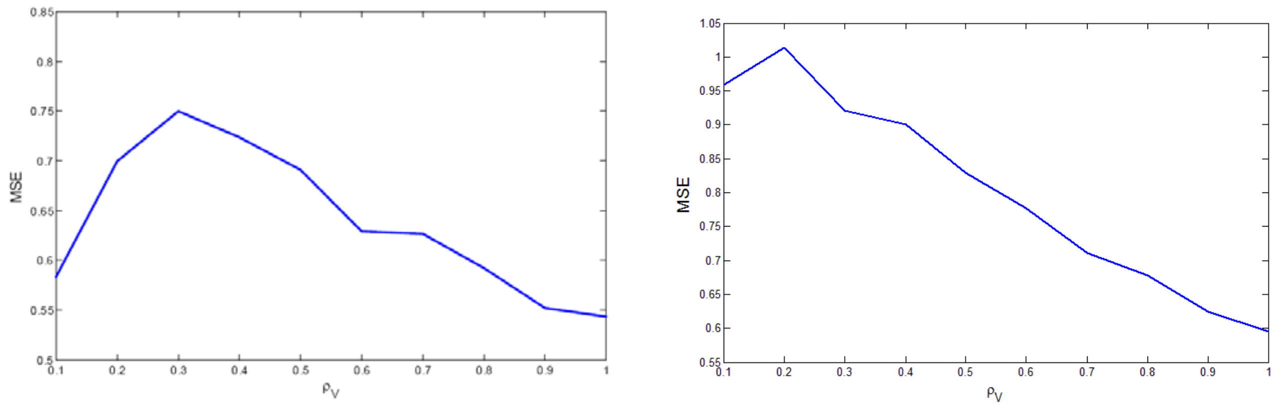
A surprising result based on Proposition 14 is that, as the homophily effect δ increases, the region in which greater precision of the public information is detrimental to the prediction market accuracy becomes smaller. In other words, increased precision of public information is less likely to be detrimental to the prediction market performance when the homophily effect is stronger. The intuition is as follows: If the role of public information precision is relatively important compared with the role of private information precision, greater public information prcesion is beneficial to the prediction market performance; otherwise, it is detrimental to the performance (a main analytical result that is robust in all our different cases). For instance, if public information precision is high, then the role of public information is relatively important, and greater public information precision is more likely to be beneficial. However, if private information precision is high, then the role of public information is relatively unimportant, and greater public information precision is more likely to be detrimental. If homophily is more signifiacant, it will reduce the informational value of friends' signals, and the role of private/public information precision will become smaller/larger because a focal participant will receive less valuable information from her friends. Therefore, greater public information precision is less likely to be detrimental.

4.3 Numerical Simulations on More Complicated Social Networks

We examine prediction markets in which participants are embedded in more complicated social networks, including the Gilbert graph, the Erdos–Renyi random graph, the "small world" graph, and the preferential attachment graph. We use CONTEST, a network toolbox for MATLAB, to simulate the aforementioned random graphs (Taylor and Higham 2009).

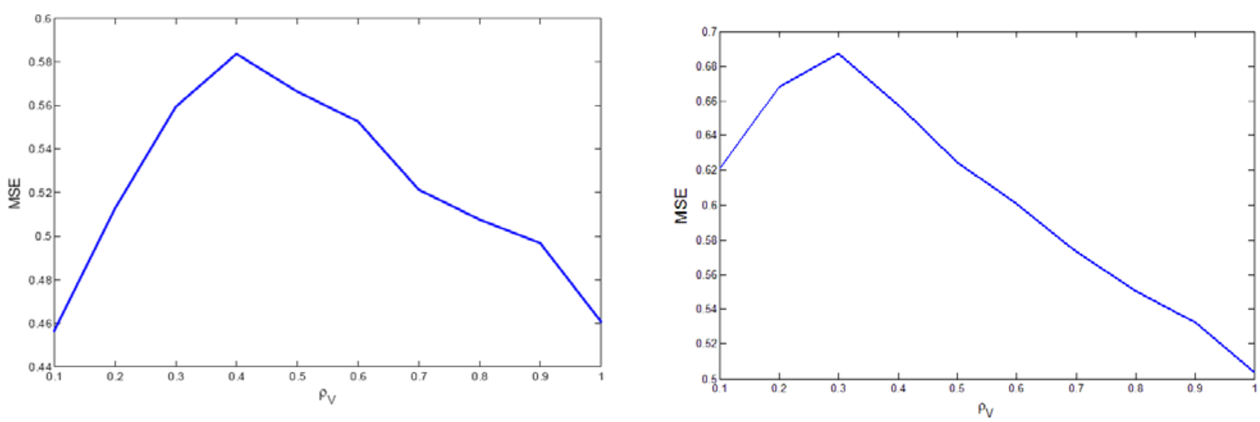
The Gilbert graph and the Erdos–Renyi random graph are classical random graph models. In Gilbert's model (Gilbert 1959), a link between two prediction market participants is formed with an independent probability p. We set the parameter values $n = 50$, $V_0 = 10$, $\rho_\varepsilon = 0.1$, and $p = 0.05$, and run the simulation 10,000 times to calculate the MSE in the prediction market. The results are robust for other parameter values. Figures 4(a) and (b) show the effect of the precision of public information on the MSE in a prediction market with a Gilbert network under no homophily/homophily (the homophily correlation coefficent is 0.2). The results are consistent in both no homophily and homophily cases. When the public information is noisy, increased precision of public information is detrimental to the prediction market performance. When public information is precise, increased precision of public information is beneficial to the prediction market performance.

In the Erdos–Renyi model (Erdos and Renyi 1960), the number of links, m, in the network is specified. We then select uniformly at random from the set of all social networks containing $n = 50$ participants and m links. We follow Taylor and Higham (2009) and set m to be the smallest integer bigger than $(n\log n)/2$. The result is similar and shown in Figure 5.



(a) No homophily (b) Homophily

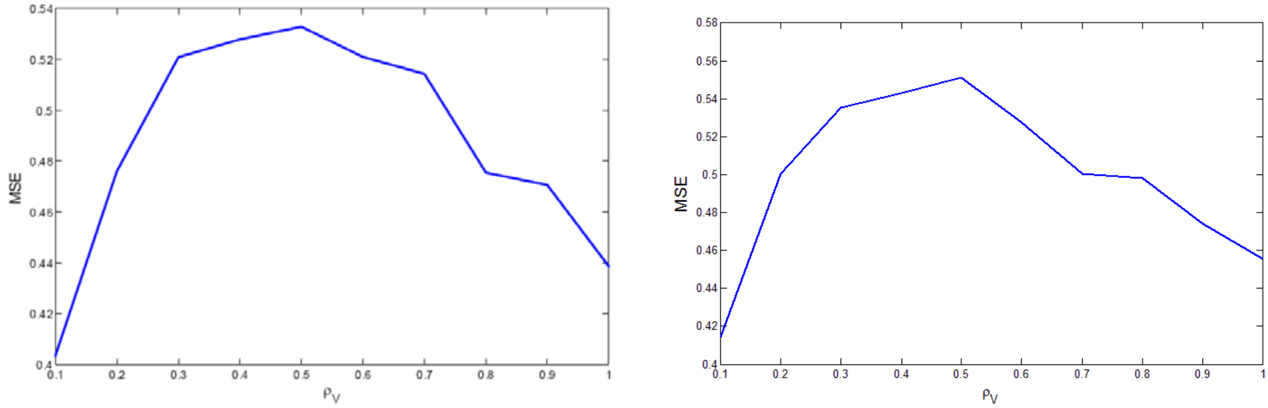
Figure 4. The Effect of the Precision of Public Information on the MSE in a Gilbert Network, $n = 50$, $V_0 = 10$, $\rho_\varepsilon = 0.1$, and $p = 0.05$.



(a) No homophily (b) Homophily

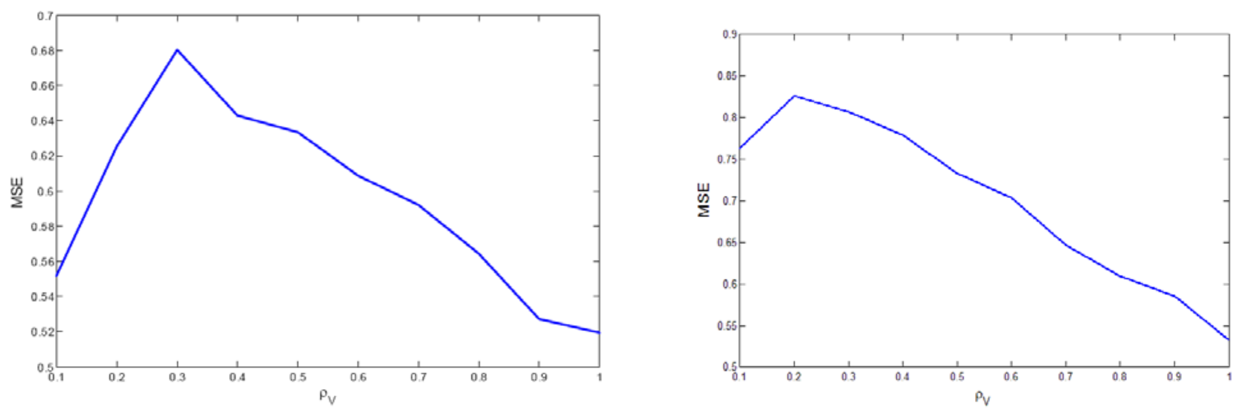
Figure 5. The Effect of the Precision of Public Information on the MSE in a Erdos–Renyi Network, $n = 50$, $V_0 = 10$, and $\rho_\varepsilon = 0.1$.

Motivated by the fact that many real-world networks have a small average shortest path length, Watts and Strogatz (1998) proposed a "small-world" network in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of steps. Following Taylor and Higham (2009), the Watts-Strogatz model begins with a k-nearest neighbor ring. Then, each participant is considered independently in turn. With a fixed probability p, a participant is given an extra link connecting it to a participant chosen uniformly at random across the network. We choose the default parameter values in Taylor and Higham (2009): $k = 2$ and $p = 0.1$. The result is shown in Figure 6.



(a) No homophily (b) Homophily

Figure 6. The Effect of the Precision of Public Information on the MSE in in a Small-World Network, $n = 50$, $V_0 = 10$, and $\rho_\varepsilon = 0.1$.



(a) No homophily (b) Homophily

Figure 7. The Effect of the Precision of Public Information on the MSE in in a Preferential Attachment Network, $n = 50$, $V_0 = 10$, and $\rho_\varepsilon = 0.1$.

The preferential attachment graph is a scale-free network that has a power-law degree distribution (Barabasi and Albert 1999). Scale-free networks are widely observed in reality. In Barabasi and Albert's model, the graph grows until n participants haven been created. Each new participant is given d links on arrival. These new connections are not chosen uniformly—the new links to an existing participant with a probability that is proportional to the current degree of that participant. In this way, well-connected participants tend to become even better connected as the graph evolves. We follow Taylor and Higham (2009) and set $d = 2$. We observe that the result in Figure 7 is similar. To provide a benchmark, we also depict the non-networked case using the same parameter values in Figure D.1, which can be found in online Appendix D.

5 Managerial Implications

The advancement of social media technologies has provided an unprecedented opportunity for corporate prediction market designers to facilitate information communications within organizations and to improve the prediction market performance. Our analytical results have the following implications and guidance for corporate prediction market design.

When should a corporate manager disclose more precise public information? Although increased precision of the public information is always beneficial to individual prediction market participants, it can be detrimental to prediction market performance as a whole when the public information is relatively noisy. When corporate prediction market designers choose the extent of public information disclosure, they need to know the level of public information precision relative to private information precision. If the private information is relatively precise, the corporation may want to hide the public information as much as possible. However, if the public information is relatively precise, the corporation may want to disclose the public information as much as possible.[20]

When should a corporate manager encourage social interactions among prediction market participants? Our model shows that social interactions among prediction market participants can improve the prediction market performance. The social network, however, has a side effect. As the level of social interaction increases, increased precision of public information may be more likely to be detrimental under some market conditions. Corporate prediction market designers should consider the pros and cons of embedding social networks in a prediction market. If increasing the level of social interactions is beneficial, managers can (i) encourage employees to be involved in multiple projects throughout the company and to become effective information hubs, and (ii) promote social networking

[20] A company can evaluate and estimate the relative level of public information precision on a case-by-case basis. Suppose that we have two corporate prediction markets: One focuses on predicting the future sales of an existing product, and the other focuses on predicting the future sales of a new product. The public information should be more precise in the first prediction market than in the second prediction market because the past sales information of the existing product is publicly available within the company and can be used to predict future sales. On the other hand, participants have less precise public information to predict future sales of new products.

among employees using Facebook, Twitter, LinkedIn, or the in-house corporate network.[21] Managers can directly measure the actual information flow among employees, such as employees' discussions, in the internal social media platform or infer the information flow through monitoring correlated prediction market trading behavior among corporate employees.[22]

More generally, our analytical model reminds corporate managers that the prediction market is not a panacea for all decision-making problems. It is true that corporate hierarchy can cause individual employees to overweight the existing public information. However, a prediction market may lead to a similar problem: The information-aggregation mechanism places a larger-than-efficient weight on the public information. As a result, increased precision of public information can be detrimental. The bottom line is that managers should be fully aware of the market conditions in which disclosing more precise public information is detrimental.

6 Conclusions

Introducing corporate prediction markets has become a popular way for companies to improve business decision making. In the present study, we examine the roles of information precision and social interactions in corporate prediction markets. The wisdom of crowds hypothesis states that existing prediction market prices always incorporate and reflect all relevant information of individuals. However, this hypothesis considers only the aggregation of diverse private information, leaving out the role of public information that is available to all participants. In our analytical model, we find that the prediction market mechanism places a larger than efficient weight on the public information. If a social network is embedded, information sharing among participants may help correct this inefficiency. Therefore, the integration of prediction markets with a social network is not only theoretically interesting (a social

[21] A designer of Google's prediction markets (GPM) pointed out that Google "should make the trading more social… should build in more social features and personalization into GPM (Coles et al. 2007, pp. 2, 14). To facilitate information exchange among participants, companies may pack them in tight in the working environment (such as sharing office rooms and providing public coffee areas), so that they can share information. Companies can also provide internal social media platforms for prediction market participants to communicate with each other. Montgomery et al. (2013) documented that Ford employees participating in the prediction market saw an internal social platform as an outlet for expressing their opinions about predictions and demonstrated a strong desire to write comments attempting to convince each other of their positions.

[22] If trading positions of two employees are highly correlated over time, it suggests an information flow between them. Using the past trading behavior data, a company can have a better idea about the actual information flows within the organization.

context is often neglected in the modeling of prediction markets), but also practically important. Our analysis should serve as a guide for corporate managers when they design internal prediction markets.

There are several possible extensions to our research. First, in our model, we assume that participants are able to observe their friends' private signals without information loss. It would be interesting to examine the information loss or bias caused by communication barriers or strategic issues (Lin, Geng, and Whinston 2005; Chen, Xu, and Whinston 2011). Second, one avenue of extending our model is to incorporate "semipublic" information that is available only to a specific group of participants in a heterogenous network. In the present paper, we distinguish two extreme types of information: private information that is received by single individuals only and public information that is available to all participants. In future research, one may allow for intermediate degrees of publicity: information that is common knowledge to only a fraction of all participants. For instance, in a company, the semipublic information in the marketing department might be different from that in the engineering department. Another potential avenue for further research is to study how social network structures affect the role of public information precision in prediction markets.

References

Abaramowicz, M., Henderson, M. T. (2007). Prediction Markets for Corporate Governance. *Notre Dame Law Review*, 82(4), 1343-1414.

Aid, R., Chemla, G., Porchet, A., Touzi, N. (2011). Hedging and Vertical Integration in Electricity Markets. *Management Science* 57(8), 1438-1452.

Angeletos, G. M., Pavan, A. (2007). Efficient Use of Information and Social Value of Information. *Econometrica* 75(4), 1103-1142.

Aral, S., Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9), 1623-1639.

Ba, S., Stallaert, J., Whinston, A. B. (2001). Optimal Investment in Knowledge within a Firm using a Market Mechanism. *Management Science*, 47(9), 1203-1219.

Banerjee, S., Kaniel, R., Kremer, I. (2009). Price Drift as an Outcome of Differences in Higher-Order Beliefs. *Review of Financial Studies*, 22(9), 3707-3734.

Banerjee, S., Kremer, I. (2010). Disagreement and learning: Dynamic patterns of trade. *Journal of Finance*, 65(4), 1269-1302.

Bapna, R., Umyarov, A. (2015). Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. *Management Science,* 61(8), 1902-1920.

Barabasi, A. L., Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509-512.

Berg, J. E., Rietz, T. A. (2003). Prediction markets as decision support systems. *Information Systems Frontiers*, 5(1), 79-93.

Berg, J., R. Forsythe, F. Nelson, T. Rietz. (2008). Results from a Dozen Years of Election Futures Markets Research. Plott, C. R., V. L. Smitt. *Handbook of Experimental Economics Results*, vol. 1, chap. 80, 742-751.

Berg, J. E., G. R. Neumann, T. A. Rietz. (2009). Searching for Google's Value: Using Prediction Markets to Forecast Market Capitalization Prior to an Initial Public Offering. *Management Science,* 55(3) 348-361.

Bloom, N., Liang, J., Roberts, J., Ying, Z. J. (2015). Does Working From Home Work? Evidence From A Chinese Experiment. *Quarterly Journal of Economics*, 130(1) 165-218.

Bloomfield, R. J., Tayler, W. B., & Zhou, F. H. (2009). Momentum, reversal, and uninformed traders in laboratory markets. *Journal of Finance*, 64(6), 2535-2558.

Bichler, M., Gupta, A., Ketter, W. (2010). Designing Smart Markets. *Information Systems Research*, 21(4), 688-699.

Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact?. *Journal of Finance*, 40(3), 793-805.

Cao, H. H., Ou-Yang, H. (2009). Differences of opinion of public information and speculative trading in stocks and options. *Review of Financial Studies*, 22(1), 299-335.

Cespa, G., Vives, X. (2015). The Beauty Contest and Short-Term Trading. *Journal of Finance,* forthcoming.

Chen, J., Xu, H., Whinston, A. B. (2011). Moderated Online Communities and Quality of User-Generated Content. *Journal of Management Information Systems*, 28(2), 237-268.

Chen, K., C. Plott. (2002). Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem. California Institute of Technology Social Science, Working Paper No. 1131.

Chen, Q., Jiang, W. (2006). Analysts' Weighting of Private and Public Information. *Review of Financial Studies*, 19(1), 319-355.

Coase, R. H. (1937). The Nature of the Firm. *Economica*, 4(16), 386-405.

Cohen, L., Frazzini, A., Malloy, C. (2008). The Small World of Investing: Board Connections and Mutual Fund Returns. *Journal of Political Economy*, 116(5), 951-979.

Coles, P. A., Lakhani, K. R., McAfee, A. (2007). Prediction markets at Google. Harvard Business School Case 9-607-088.

Colla, P., Mele, A. (2010). Information Linkages and Correlated Trading. *Review of Financial Studies*, 23(1), 203-246.

Corgnet, B., DeSantis, M., Porter, D. (2015). Revisiting Information Aggregation in Asset Markets: Reflective Learning & Market Efficiency. Working paper, Chapman University, Economic Science Institute, available at http://econpapers.repec.org/paper/chuwpaper/15-15.htm.

Coval, J. D., T. J. Moskowitz. (2001). The Geography of Investment: Informed Trading and Asset Prices. *Journal of Political Economy* 109(4) 811-841.

Cowgill, B., Wolfers, J., Zitzewitz, E. (2009). Using Prediction Markets to Track Information Flows: Evidence from Google. In Auctions, Market Mechanisms and Their Applications: First International ICST Conference, AMMA 2009, Boston, MA, USA, May 8-9, 2009, Revised Selected Papers (Vol. 14, p. 3). Springer Science & Business Media.

Cowgill, B., Zitzewitz, E. (2014). Corporate Prediction Markets: Evidence from Google, Ford, and Firm X1. *Review of Economic Studies*, forthcoming.

Davis-Stober, C. P., Budescu, D. V., Broomell, S. B., Dana, J. (2015). The Composition of Optimally Wise Crowds. *Decision Analysis*, forthcoming.

Erdos, P., Renyi, A. (1960). On the Evolution of Random Graphs. *Publications of the Mathematical*

Institute of the Hungarian Academy of Sciences, 5, 17-61.

Fang, F., M. Stinchcombe, A. B. Whinston. (2007). Putting Your Money Where Your Mouth Is – Betting Mechanism Design for Better Prediction. *Review of Network Economics* 6(3) 214-238.

Fang, F., Stinchcombe, M., A. B. Whinston. (2010). Proper Scoring Rules with Arbitrary Value Functions. *Journal of Mathematical Economics* 46(6) 1200-1210.

Foutz, N. Z., Jank, W. (2010). Prerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Marketing Science*, 29(3), 568-579.

Gilbert, E. N. (1959). Random Graphs. *Annals of Mathematical Statistics*, 30(4), 1141-1144.

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91(3), 481-510.

Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *Journal of Finance*, 31(2), 573-585.

Grossman, S. (1978). Further results on the informational efficiency of competitive stock markets. *Journal of Economic Theory*, 18(1), 81-101.

Grossman, S. J., Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 393-408.

Golub, B., M. Jackson. (2010). Naive Learning in Social Networks and the Wisdom of Crowds. American Economic Journal: Microeconomics 2(1) 112-149.

Gu, B., Konana, P., Raghunathan, R., Chen, H. M. (2014). The Allure of Homophily in Social Media: Evidence from Investor Responses on Virtual Communities. *Information Systems Research*, 25(3), 604-617.

Guo, Z., F. Fang, A. B. Whinston. (2006). Supply Chain Information Sharing in a Macro Prediction Market. *Decision Support Systems* 42(3) 1944-1958.

Han, B., Yang, L. (2013). Social Networks, Information Acquisition, and Asset Prices. *Management Science*, 59(6), 1444-1457.

Hansen, M. T. (2002). Knowledge networks: Explaining effective knowledge sharing in multiunit companies. *Organization Science*, 13(3), 232-248.

Harris, M., Raviv, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies*, 6(3), 473-506.

Harrison, J. M., and Kreps, D. M. (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *Quarterly Journal of Economics*, 92(2), 323-336.

Healy, P. J., S. Linardi, J. R. Lowery, J. O. Ledyard. (2010). Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders. *Management Science* 56(11) 1977-1996.

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, 22(3), 477-498.

Hirshleifer, D., Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1), 337-386.

Hong, H., Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54(6), 2143-2184.

Hong, H., Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *Review of Financial Studies*, 16(2), 487-525.

Hopman, J. W. (2007). Using forecasting markets to manage demand risk. *Intel Technology Journal*, 11(2), 127-135.

Jackson, M. O. (2008). Social and Economic Networks, Princeton University Press, Princeton, NJ.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2-3), 95-101.

Jian, L., R. Sami. (2012). Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution. *Management Science* 58(1) 123–140.

Jiang, Z., Mookerjee, V. S., Sarkar, S. (2005). Lying on the web: implications for expert systems redesign. *Information Systems Research*, 16(2), 131-148.

Jiang, Y., Guo, H. (2015). Design of Consumer Review Systems and Product Pricing. *Information Systems Research*, forthcoming.

Kandel, E., Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4), 831-872.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449-1475.

Ketter, W., Peters, M., Collins, J., Gupta, A. (2015). Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data and Analytics. *MIS Quarterly*, forthcoming.

Ketter, W., Peters, M., Collins, J., Gupta, A. (2016). A Multi-agent Competitive Gaming Platform to address Societal Challenges. *MIS Quarterly*, Forthcoming.

Keuschnigg, M., Ganser, C. (2016). Crowd Wisdom Relies on Agents' Ability in Small Groups with a Voting Aggregation Rule. *Management Science*, forthcoming.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315-1335.

Lamberson, P. J., Page, S. E. (2012). Optimal Forecasting Groups. *Management Science*, 58(4), 805-810.

Leigh, A., Wolfers, J. (2007). Prediction markets for business and public policy. *Melbourne Review: A Journal of Business and Public Policy*, 3(1), 7-15.

Levy, H., Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *American Economic Review*, 308-317.

Lin, L., Geng, X., Whinston, A. B. (2005). A Sender-Receiver Framework for Knowledge Transfer. *MIS Quarterly*, 29(2), 197-219.

Liu, D., Geng, X., Whinston, A. B. (2007). Optimal Design of Consumer Contests. *Journal of Marketing*, 71(4), 140-155.

Lorenz, J., Rauhut, H., Schweitzer, F., Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020-9025.

Lovell, M. C. (1986). Tests of the rational expectations hypothesis. *American Economic Review*, 76(1), 110-124.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2), 383-417.

Mello, J. A. (2014). Strategic Human Resource Management. Cengage Learning, Stamford, CT.

McHugh, P., and Jackson, A. (2012). Prediction market accuracy: the impact of size, incentives, context and interpretation. *Journal of Prediction Markets*, 6(2), 22-46.

Montgomery, T. A., Stieg, P. M., Cavaretta, M. J., Moraal, P. E. (2013). Experience from hosting a corporate prediction market: benefits beyond the forecasts. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1384-1392). ACM.

Morris, S., H. S. Shin. (2002). Social Value of Public Information. *American Economic Review* 92(5), 1521–1534.

O'hara, M. (1995). Market microstructure theory. Cambridge, MA: Blackwell.

Ozsoylev, H. N., Walden, J. (2011). Asset pricing in large information networks. *Journal of Economic Theory*, 146(6), 2252-2280.

Parise, S., Whelan, E., Todd, S. (2015). How Twitter Users Can Generate Better Ideas. *MIT Sloan Management Review*, 56(4), 21.

Peng, L. (2005). Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis*, 40(2), 307-329.

Prendergast, C. (1993). A Theory of Yes Men. *American Economic Review* 83(4) 757-70.

Qiu, L., Rui, H., Whinston, A. B. (2014a). Effects of Social Networks on Prediction Markets: Examination in a Controlled Experiment. *Journal of Management Information Systems*, 30(4), 235-268.

Qiu, L., Rui, H., Whinston, A. B. (2014b). The Impact of social network structures on prediction market accuracy in the presence of insider information. *Journal of Management Information Systems*, 31(1), 145-172.

Scheinkman, J. A., Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of Political Economy*, 111(6), 1183-1220.

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?. *American Economic Review*, 71(3), 421-436.

Stasser, G., Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6), 1467.

Stasser, G., Titus, W. (2003). Hidden profiles: A brief history. *Psychological Inquiry*, 14(3-4), 304-313.

Surowiecki, J. (2004). The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations, Doubleday, New York, NY.

Sunstein, C. R. (2005). Group Judgements: Statistical Mmeans, Deliberation, and Information Markets. *NYU Law Review*, 80, 962-1895.

Taylor, A., Higham, D. J. (2009). CONTEST: A Controllable Test Matrix Toolbox for MATLAB. *ACM Transactions on Mathematical Software* (TOMS), 35(4), 26.

Tirole, J. (1982). On the Possibility of Speculation under Rational Expectations. *Econometrica*, 50(5), 1163-1181.

Van Bruggen, G. H., Spann, M., Lilien, G. L., Skiera, B. (2010). Prediction markets as institutional forecasting support systems. *Decision Support Systems*, 49(4), 404-416.

Vives, X. (1993). How fast do rational agents learn?. *Review of Economic Studies*, 60(2), 329-347.

Watts, D. J., Strogatz, S. H. (1998). Collective Dynamics of "Small-World" Networks. *Nature*, 393(6684), 440-442.

Whelan, E., Parise, S., De Valk, J., Aalbers, R. (2011). Creating employee networks that deliver open innovation. *MIT Sloan Management Review*, 53(1), 37-44.

"Hidden Profiles" in Corporate Prediction Markets: The Impact of Public Information Precision and Social Interactions (Online Appendix)

Online Appendix A: Proof

Proof of Proposition 1

Proof. Each participant's demand for the security is given by equation 4. We solve the equilibrium prediction market price P^* by plugging equation 4 into the market clearing condition, $\sum_{i=1}^{n} x_i^* = 0$. Then we can obtain the equilibrium demand x_i^*.

Proof of Proposition 2

Proof. From equation 7, it is obvious that $\text{MSE}(P^*)$ decreases with n. We can also obtain:

$$\frac{\partial}{\partial \rho_\varepsilon}\text{MSE}(P^*) = \frac{1}{(\rho_\varepsilon+\rho_V)^3}\left[-\frac{\rho_\varepsilon}{n} + \left(\frac{1}{n} - 2\right)\rho_V\right] < 0,$$

and

$$\frac{\partial}{\partial \rho_V}\text{MSE}(P^*) = \frac{1}{(\rho_\varepsilon+\rho_V)^3}\left[\left(\frac{n-2}{n}\right)\rho_\varepsilon - \rho_V\right].$$

Therefore, if $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{n-2}{n}$, $\frac{\partial}{\partial \rho_V}\text{MSE}(P^*) \geq 0$; if $\frac{\rho_V}{\rho_\varepsilon} > \frac{n-2}{n}$, $\frac{\partial}{\partial \rho_V}\text{MSE}(P^*) < 0$.

Proof of Proposition 3

Proof. The marginal line $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n}$ determines the range of two regions (whether or not increased precision of public information is detrimental), and $\frac{n-2}{n}$ increases with n. Therefore, as n increases, the region, $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{n-2}{n}$, becomes larger, and the region, $\frac{\rho_V}{\rho_\varepsilon} > \frac{n-2}{n}$, shrinks.

Proof of Proposition 4

Proof. Each participant's demand for the security is given by equation $x_i^* = \frac{\mathbf{E}[V|I_i]-P}{2\gamma\mathbf{Var}[V|I_i]}$. For a DO trader $i \in C_{DO}$:

$$\mathbf{E}[V|I_i] = \mathbf{E}[V|S_i] = \frac{\rho_V}{\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}S_i,$$

$$\mathbf{Var}[V|I_i] = \mathbf{Var}[V|S_i] = 1/(\rho_\varepsilon + \rho_V).$$

For an REE trader $i \in C_{REE}$:

$$\mathbf{E}[V|I_i] = \mathbf{E}[V|S_i, P^*]$$

$$= \frac{\rho_V}{\left(\frac{nb^2}{c^2}+1\right)\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{\left(\frac{nb^2}{c^2}+1\right)\rho_\varepsilon+\rho_V}S_i + \frac{\frac{nb^2}{c^2}\rho_\varepsilon}{\left(\frac{nb^2}{c^2}+1\right)\rho_\varepsilon+\rho_V}\left(\frac{P^*-a}{b}\right),$$

$$\mathbf{Var}[V|I_i] = \mathbf{Var}[V|S_i, P^*] = 1/\left[\left(\frac{nb^2}{c^2}+1\right)\rho_\varepsilon + \rho_V\right].$$

We solve the equilibrium prediction market price P^* by plugging these equations into the market clearing condition, $\sum_{i\in C_{DO}} x_i^* + \sum_{j\in C_{REE}} x_j^* = 0.$ Then we compare the solution from the market clearing condition with the initial conjecture:

$$P^* = a + bV + c\bar{\varepsilon},$$

and determine the coefficents *a*, *b*, and *c*.

Proof of Proposition 5

Proof. When $m \geq 1$,

$$\frac{\partial}{\partial m}\text{MSE}(P^*) = \frac{2\rho_\varepsilon\rho_V}{[(n+1-m)\rho_\varepsilon+\rho_V]^3} - \frac{2(n+1-m)\rho_\varepsilon}{[(n+1-m)\rho_\varepsilon+\rho_V]^2} + \frac{2(n+1-m)^2\rho_\varepsilon^2}{n[(n+1-m)\rho_\varepsilon+\rho_V]^3} = \frac{2(m-1)\rho_\varepsilon\rho_V}{n[(n+1-m)\rho_\varepsilon+\rho_V]^3} \geq 0.$$

Proof of Proposition 6

Proof. $\frac{\partial}{\partial n}\text{MSE}(P^*) = -\frac{2\rho_\varepsilon\rho_V}{[(n+1-m)\rho_\varepsilon+\rho_V]^3} + \frac{2(n+1-m)\rho_\varepsilon}{n[(n+1-m)\rho_\varepsilon+\rho_V]^2} - \frac{(n+1-m)^2\rho_\varepsilon}{n^2[(n+1-m)\rho_\varepsilon+\rho_V]^2} -$

$\frac{2(n+1-m)^2\rho_\varepsilon\rho_V}{n[(n+1-m)\rho_\varepsilon+\rho_V]^3} = \frac{-2(m-1)n\rho_\varepsilon\rho_V-(n+1-m)^2[(n+1-m)\rho_\varepsilon+\rho_V]\rho_\varepsilon}{n^2[(n+1-m)\rho_\varepsilon+\rho_V]^3} < 0.$

$\frac{\partial}{\partial\rho_\varepsilon}\text{MSE}(P^*) = -\frac{2(n+1-m)\rho_V}{[(n+1-m)\rho_\varepsilon+\rho_V]^3} + \frac{(n+1-m)^2}{n[(n+1-m)\rho_\varepsilon+\rho_V]^2} - \frac{2(n+1-m)^3\rho_\varepsilon}{n[(n+1-m)\rho_\varepsilon+\rho_V]^3} =$

$\frac{-(n+1-m)(m+n-1)\rho_V-(n+1-m)^3\rho_\varepsilon}{n^2[(n+1-m)\rho_\varepsilon+\rho_V]^3} < 0.$

$\frac{\partial}{\partial\rho_V}\text{MSE}(P^*) = \frac{1}{[(n+1-m)\rho_\varepsilon+\rho_V]^2} - \frac{2\rho_V}{[(n+1-m)\rho_\varepsilon+\rho_V]^2} - \frac{2(n+1-m)^2\rho_\varepsilon}{n[(n+1-m)\rho_\varepsilon+\rho_V]^3} =$

$\frac{(n+1-m)(2m-n-2)\rho_\varepsilon-n\rho_V}{n^2[(n+1-m)\rho_\varepsilon+\rho_V]^3}.$

If $m \le \frac{n+2}{2}$, $\frac{\partial}{\partial\rho_V}\text{MSE}(P^*) < 0$. If $m > \frac{n+2}{2}$ and $\frac{\rho_V}{\rho_\varepsilon} \le \frac{(2m-2-n)(n+1-m)}{n}$, $\frac{\partial}{\partial\rho_V}\text{MSE}(P^*) \ge 0$; if $\frac{\rho_V}{\rho_\varepsilon} > \frac{(2m-2-n)(n+1-m)}{n}$, $\frac{\partial}{\partial\rho_V}\text{MSE}(P^*) < 0$.

Proof of Proposition 7

Proof. We plug $x_i^* = \frac{\mathbf{E}[V|I_i]-P}{2\gamma\mathbf{Var}[V|I_i]}$ into the market clearing condition and obtain:

$$P^* = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[V|I_i] = \frac{\rho_V}{(k+1)\rho_\varepsilon+\rho_V}V_0 + \frac{(k+1)\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}V + \frac{(k+1)\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}\bar{\varepsilon}.$$

Then,

$$x_i^* = \frac{\mathbf{E}[V|I_i]-P^*}{2\gamma\mathbf{Var}[V|I_i]} = \frac{\rho_\varepsilon}{2\gamma}\left[\varepsilon_i + \sum_{j\in N_i(g)} \varepsilon_j - (k+1)\bar{\varepsilon}\right].$$

Proof of Proposition 8

Proof. From equations 7 and 11, we can obtain the difference between the MSE in a prediction market without social networks and the MSE in a prediction market with a regular social network:

$$\frac{\rho_V}{(\rho_\varepsilon+\rho_V)^2}+\frac{\rho_\varepsilon}{n(\rho_\varepsilon+\rho_V)^2}-\frac{\rho_V}{[(k+1)\rho_\varepsilon+\rho_V]^2}-\frac{\rho_\varepsilon(k+1)^2}{n[(k+1)\rho_\varepsilon+\rho_V]^2}\geq 0,$$

where the equality holds when $k=0$.

Proof of Proposition 9

Proof. From equation 11, we can obtain:

$$\frac{\partial}{\partial k}\mathrm{MSE}(P^*)=\frac{1}{[(k+1)\rho_\varepsilon+\rho_V]^3}\left[-2\rho_\varepsilon\rho_V+\frac{2}{n}\rho_\varepsilon\rho_V(k+1)\right]\leq 0.$$

The inequality comes from the fact that $k+1\leq n$ in a regular network.

Proof of Proposition 10

Proof. From equation 11, it is obvious that $\mathrm{MSE}(P^*)$ decreases with n. From equation 11, we can also obtain:

$$\frac{\partial}{\partial \rho_\varepsilon}\mathrm{MSE}(P^*)$$

$$=\frac{1}{[(k+1)\rho_\varepsilon+\rho_V]^3}\left[-\frac{\rho_\varepsilon(k+1)^3}{n}+\left(\frac{k+1}{n}-2\right)\rho_V(k+1)\right]<0,$$

and

$$\frac{\partial}{\partial \rho_V}\mathrm{MSE}(P^*)$$

$$=\frac{1}{[(k+1)\rho_\varepsilon+\rho_V]^3}\left[\left(\frac{n-2(k+1)}{n}\right)(k+1)\rho_\varepsilon-\rho_V\right].$$

Therefore, the result follows.

Proof of Proposition 11

Proof. The marginal line $\frac{\rho_V}{\rho_\varepsilon}=(k+1)\left[\frac{n-2(k+1)}{n}\right]$ determines the range of two regions (whether or not increased precision of public information is detrimental). The right hand side $(k+$

1) $\left[\frac{n-2(k+1)}{n}\right]$ increases with n, increases with k if $n \geq 4(k+1)$, and decreases with k if $n < 4(k+1)$.

Proof of Proposition 12

Proof. Each participant's demand for the security is given by equation $x_i^* = \frac{\mathbf{E}[V|I_i]-P}{2\gamma\mathbf{Var}[V|I_i]}$. In the homophily case,

$$\mathbf{E}[V|I_i] = \frac{\rho_V}{\frac{2}{1+\delta}\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{\frac{2}{1+\delta}\rho_\varepsilon+\rho_V}\frac{1}{1+\delta}S_i + \frac{\rho_\varepsilon}{\frac{2}{1+\delta}\rho_\varepsilon+\rho_V}\frac{1}{1+\delta}S_j,\ \mathbf{Var}[V|I_i] = 1/\left[\frac{2}{1+\delta}\rho_\varepsilon + \rho_V\right].$$

We solve the equlibrium price by using the market clearing condition $\sum_{i=1}^n x_i^* = 0$.

Proof of Proposition 13

Proof. $\frac{\partial}{\partial\delta}\text{MSE}(P^*) = \frac{4\rho_V\rho_\varepsilon}{(1+\delta)^2(\rho_V+\frac{2\rho_\varepsilon}{1+\delta})^3} + \frac{16\rho_\varepsilon^2}{n(1+\delta)^3(\rho_V+\frac{2\rho_\varepsilon}{1+\delta})^3} - \frac{4\rho_\varepsilon}{n(1+\delta)^2(\rho_V+\frac{2\rho_\varepsilon}{1+\delta})^2} = \frac{4\rho_\varepsilon[(-1+n)(1+\delta)\rho_V+2\rho_\varepsilon]}{n[(1+\delta)\rho_V+2\rho_\varepsilon]^3} > 0.$

Proof of Proposition 14

Proof. $\frac{\partial}{\partial n}\text{MSE}(P^*) = -\frac{4\rho_\varepsilon}{n^2(1+\delta)\left(\rho_V+\frac{2\rho_\varepsilon}{1+\delta}\right)^2} < 0.$

$\frac{\partial}{\partial\rho_\varepsilon}\text{MSE}(P^*) = -\frac{4\rho_V}{(1+\delta)\left(\rho_V+\frac{2\rho_\varepsilon}{1+\delta}\right)^3} - \frac{16\rho_\varepsilon}{n(1+\delta)^2\left(\rho_V+\frac{2\rho_\varepsilon}{1+\delta}\right)^3} + \frac{4}{n(1+\delta)\left(\rho_V+\frac{2\rho_\varepsilon}{1+\delta}\right)^2} = -\frac{4(1+\delta)[(-1+n)(1+\delta)\rho_V+2\rho_\varepsilon]}{n[(1+\delta)\rho_V+2\rho_\varepsilon]^3} < 0.$

$\frac{\partial}{\partial\rho_V}\text{MSE}(P^*) = -\frac{2\rho_V}{(\rho_V+\frac{2\rho_\varepsilon}{1+\delta})^3} - \frac{8\rho_\varepsilon}{n(1+\delta)(\rho_V+\frac{2\rho_\varepsilon}{1+\delta})^3} + \frac{1}{(\rho_V+\frac{2\rho_\varepsilon}{1+\delta})^2} = -\frac{(1+\delta)^2[n(1+\delta)\rho_V-2(-4+n)\rho_\varepsilon]}{n[(1+\delta)\rho_V+2\rho_\varepsilon]^3}.$

If $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{2}{1+\delta}\left[\frac{n-4}{n}\right]$, $\frac{\partial}{\partial\rho_V}\text{MSE}(P^*)$ is positive; if $\frac{\rho_V}{\rho_\varepsilon} > \frac{2}{1+\delta}\left[\frac{n-4}{n}\right]$, $\frac{\partial}{\partial\rho_V}\text{MSE}(P^*)$ is negative.

Online Appendix B: A Heterogenous Social Network

In a heterogenous social network, we have two types of participants: (i) participants whose degree is 0; and (ii) participants whose degree is k. The proportions of degree 0 and degree k participants are a_0 and $1 - a_0$, respectively. Note that if $a_0 = 1$ or 0, a prediction market with a heterogenous social network will degenerate to two special cases: a non-networked prediction market ($a_0 = 1$) or a prediction market with a regular social network ($a_0 = 0$). We denote the set of degree 0 participants by D_0, and the set of degree k participants by D_k. Therefore, the set of all participants $N = D_0 \cup D_k$.

In a heterogenous social network, degree 0 and degree k participants have different information sets. The inference process of a degree 0 participant is similar to that of a participant in a non-networked prediction market. For an individual $i \in D_0$, she makes an inference using her own private signal and the common prior:

$$\mathbf{E}_0[V|I_i] = \mathbf{E}_0[V|S_i] = \frac{\rho_V}{\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}S_i,$$

$$\mathbf{Var}_0[V|I_i] = 1/(\rho_\varepsilon + \rho_V),$$

where $\mathbf{E}_0[V|I_i]$ and $\mathbf{Var}_0[V|I_i]$ are the conditional expectation and conditional variance of a degree 0 participant.

The inference process of a degree k participant is similar to that of a participant in a prediction market with a regular network. A degree k participant's information set includes her private signal, her friends' private signals (k signals), and the common prior. For an individual $j \in D_k$, she makes an inference as follows:

$$\mathbf{E}_k\left[V|I_j\right] = \frac{\rho_V}{(k+1)\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}S_i + \sum_{h\in N_j(g)} \frac{\rho_\varepsilon}{(k+1)\rho_\varepsilon+\rho_V}S_h,$$

$$\mathbf{Var}_k\left[V|I_j\right] = 1/[(k+1)\rho_\varepsilon + \rho_V],$$

where $\mathbf{E}_k[V|I_j]$ and $\mathbf{Var}_k[V|I_j]$ are the conditional expectation and conditional variance of a degree k participant. The market clearing condition is given by:

$$\sum_{i\in D_0} x_i^* + \sum_{j\in D_k} x_j^* = 0,$$

where x_i^* and x_j^* indicate the optimal positions of degree 0 and degree k participants respectively and are given as follows:

$$x_i^* = \frac{\mathbf{E}_0[V|I_i]-P}{2\gamma\mathbf{Var}_0[V|I_i]}, x_j^* = \frac{\mathbf{E}_k[V|I_j]-P}{2\gamma\mathbf{Var}_k[V|I_j]}.$$

The equilibrium is characterized in the following proposition:

Proposition B.1 (Prediction Market Equilibrium in a Heterogeneous Social Network) *In a prediction market with a heterogenous social network, the equilibrium prediction market price is given by*

$$P^* = \frac{\frac{a_0 n}{\mathbf{Var}_0[V|I_i]}}{\frac{a_0 n}{\mathbf{Var}_0[V|I_i]}+\frac{(1-a_0)n}{\mathbf{Var}_k[V|I_j]}}\sum_{i\in D_0}\frac{\mathbf{E}_0[V|I_i]}{a_0 n} + \frac{\frac{(1-a_0)n}{\mathbf{Var}_k[V|I_j]}}{\frac{a_0 n}{\mathbf{Var}_0[V|I_i]}+\frac{(1-a_0)n}{\mathbf{Var}_k[V|I_j]}}\sum_{j\in D_k}\frac{\mathbf{E}_k[V|I_j]}{(1-a_0)n}$$

$$= \delta V_0 + (1-\delta)V + \frac{\rho_\varepsilon}{[a_0+(1-a_0)(k+1)]\rho_\varepsilon+\rho_V}\left[\sum_{i\in D_0}\frac{\varepsilon_i}{n} + (k+1)\sum_{j\in D_k}\frac{\varepsilon_j}{n}\right],$$

where $\delta = \frac{\rho_V}{[a_0+(1-a_0)(k+1)]\rho_\varepsilon+\rho_V}$. *The equilibrium position for individual* $i \in D_0$ *is* $x_i^* = \frac{\boldsymbol{E}_0[V|I_i]-P^*}{2\gamma\boldsymbol{Var}_0[V|I_i]}$, *and the equilibrium position for individual* $j \in D_k$ *is* $x_j^* = \frac{\boldsymbol{E}_k[V|I_j]-P^*}{2\gamma\boldsymbol{Var}_k[V|I_j]}$.

The market price, P^*, in a heterogenous social network is a weighted average of the individual expectations, and the weight depends on $\mathbf{Var}_0[V|I_i]$ and $\mathbf{Var}_k[V|I_j]$. In a non-networked prediction market or a prediction market with a regular network, P^* is a simple average of individual expectations and is independent of $\mathbf{Var}[V|I_i]$. This is because in these two cases, $\mathbf{Var}[V|I_i]$ is the same across participants and cancels in the market clearing condition. However, in a heterogenous social network, $\mathbf{Var}_0[V|I_i] \neq \mathbf{Var}_k[V|I_j]$, so P^* depends on both

$\mathbf{Var}_0[V|I_i]$ and $\mathbf{Var}_k[V|I_j]$.

Then, we compute the MSE of P^* in a prediction market with a heterogenous social network:

$$\text{MSE}(P^*) = \mathbf{E}[(V - P^*)^2] = \frac{\rho_V + \frac{1}{n} a_0 \rho_\varepsilon + \frac{1}{n} \rho_\varepsilon (1 - a_0)(k+1)^2}{\left[[a_0 + (1 - a_0)(k+1)]\rho_\varepsilon + \rho_V\right]^2}. \quad \text{(B.1)}$$

Note that when $a_0 = 1$, the MSE in a prediction market with a heterogenous social network will be degenerated to equation 7, the MSE in a non-networked prediction market; when $a_0 = 0$, the MSE in equation B.1 will be degenerated to equation 11, the MSE in a regular network. When $n \to \infty$, the MSE in equation B.1 converges to $\frac{\rho_V}{\left[[a_0 + (1 - a_0)(k+1)]\rho_\varepsilon + \rho_V\right]^2}$. If we compare the MSEs in different cases, we obtain the following proposition:

Proposition B.2 (MSE Comparison) *When $n \to \infty$, the MSE in a non-networked prediction market is greater than the MSE in a prediction market with a heterogenous social network, and the MSE in a prediction market with a heterogenous social network is greater than the MSE in a prediction market with a regular social network.*

In the following proposition, we examine the impact of the precision of public and private information.

Proposition B.3 (Comparative Statics on MSE) *In a prediction market with a heterogenous social network, the MSE of the forecast P^* decreases with the number of prediction market participants, n, and the precision of private signals, ρ_ε. If $\frac{\rho_V}{\rho_\varepsilon} \leq \frac{n-2}{n} a_0 + (1 - a_0)(k + 1)\left[\frac{n-2(k+1)}{n}\right]$, the MSE increases with the precision of public information; if $\frac{\rho_V}{\rho_\varepsilon} > \frac{n-2}{n} a_0 + (1 - a_0)(k + 1)\left[\frac{n-2(k+1)}{n}\right]$, the MSE decreases with the precision of public information.*

Similarly, Proposition B.3 shows that in a prediction market with a heterogenous social network, increased precision of private information always enhances the prediction market accuracy, but increased precision of public information might be detrimental under some market conditions. The marginal line in a heterogenous social network, $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n} a_0 + (1 - a_0)(k + 1)\left[\frac{n-2(k+1)}{n}\right]$, is between the marginal line in a non-networked prediction market, $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n}$, and the marginal line in a regular social network, $\frac{\rho_V}{\rho_\varepsilon} = (k+1)\left[\frac{n-2(k+1)}{n}\right]$. The following numerical example illustrates the market conditions in which increased precision of public information is detrimental in a prediction market with a heterogenous social network. Figure B.1 depicts the contour lines of the MSE in a heterogenous social network when $n = 50$, $k = 9$, and $a_0 = 0.8$. The marginal line in a heterogenous social network (the solid line) is between the marginal line in a non-networked prediction market (the dashed line) and the marginal line in a regular social network (the dash-dot line). It means that Region II in a heterogenous network is larger than that in a non-networked prediction market, but smaller than that in a regular network under the chosen parameter values. The intuition is that a heterogenous network is a linear combination of a regular network and a non-networked environment. Therefore, the marginal line in a heterogenous social network, $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n} a_0 + (1 - a_0)(k + 1)\left[\frac{n-2(k+1)}{n}\right]$, is a linear combination of the two: if $a_0 = 1$, $\frac{\rho_V}{\rho_\varepsilon} = \frac{n-2}{n}$; if $a_0 = 0$, $\frac{\rho_V}{\rho_\varepsilon} = (k+1)\left[\frac{n-2(k+1)}{n}\right]$.

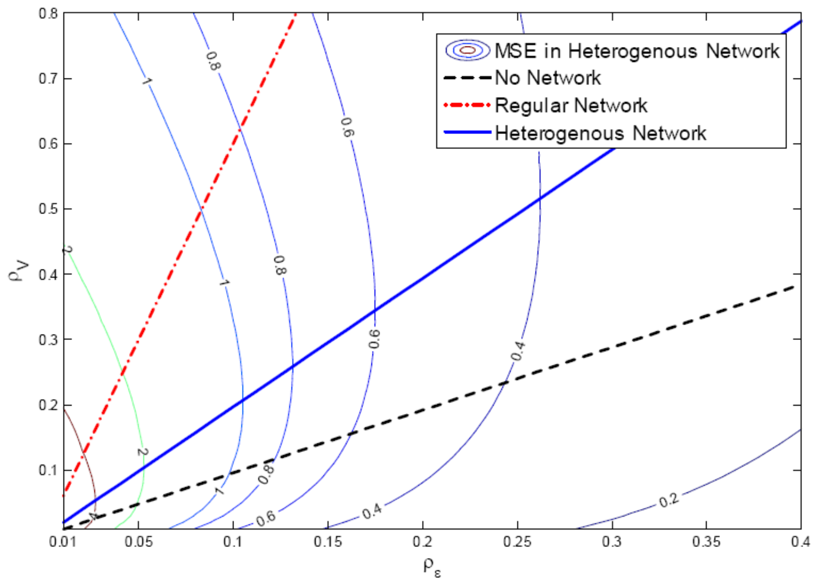


Figure B.1. The Impact of Public Information Precision on Prediction Market Performance (Heterogenous Network), $n = 50, k = 9, a_0 = 0.8$.

Online Appendix C: Selection of Prediction Market Participants

An interesting observation from Propositions 8, 9, and 10 is that a socially embedded prediction market with low precision of private information may perform as well as a non-networked prediction market with high precision of private information. The following numerical example in Figure C.1 illustrates the impacts of the precision of private information, ρ_ε, and the level of social interactions, k, on prediction market performance when $n = 50$ and $\rho_V = 0.2$. As we expected, prediction market performance increases with ρ_ε and k. In a non-networked prediction market ($k = 0$), if $\rho_\varepsilon = 0.125$, the MSE is around 2. To reach a similar level of MSE, much lower precision of private information is needed in a socially embedded prediction market with $k = 5$: $\rho_\varepsilon = 0.025$.

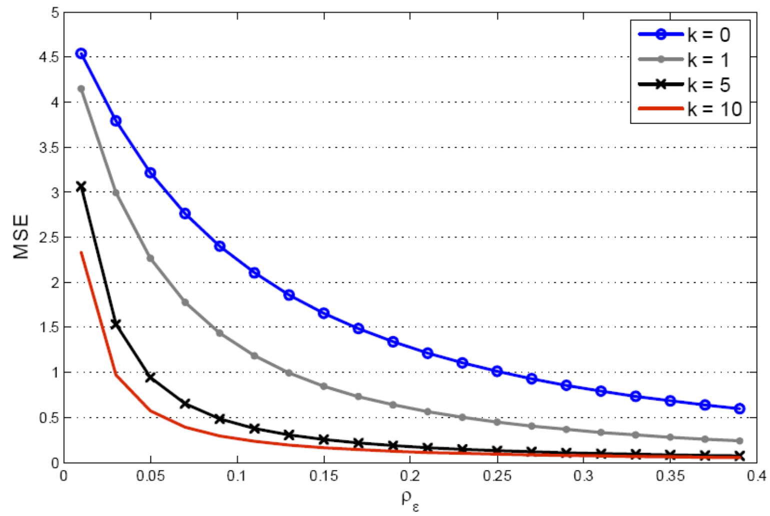


Figure C.1 The Impact of Social Interactions on Prediction Market Performance, $n = 50$, and $\rho_V = 0.2$.

A managerial implication of this result is about the selection of prediction market participants. In general, an internal employee has two types of skills: "work skills" and "social skills." In our context, the level of work skills refers to the ability to acquire precise private information (knowledge creation and information production) and is measured by ρ_ε. In contrast, the level of social skills refers to the ability to communicate and share information with colleagues (knowledge transfer and information communication) and is measured by k. Intuitively, a manager should select employees who have a high level of work skills (ρ_ε) as prediction market participants. This is also consistent with Proposition 10. However, Propositions 8 and 9 show that the level of social skills (k) also matters when we consider prediction market performance. A group of participants who have a medium level of work skills but a high level of social skills may outperform those who have a high level of work skills but a low level of social skills. Actually, Figure C.1 visually shows this implication by varying ρ_ε and k.

Online Appendix D: Additional Numerical Analysis

To provide a benchmark, in Figure D.1, we depict the non-networked case using the same paramater values as those in Section 4.3. The pattern is similar, but we have two additional observations: (i) The MSE in a non-networked prediction market is significantly greater than that in a socially embedded prediction market, which is consistent with the spirit of Proposition 5. (ii) The range of a detrimental effect of public information is smaller in a non-networked prediction market than in a socially embedded prediction market under the chosen parameter values. This is reminiscent of Proposition 11.

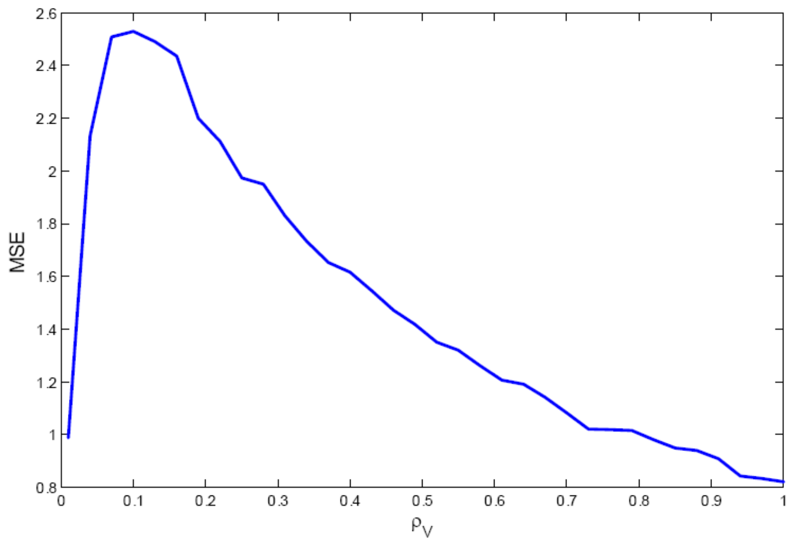


Figure D.1 The Effect of the Precision of Public Information on the MSE in a Non-Networked Prediction Market, $n = 50$, $V_0 = 10$, and $\rho_\varepsilon = 0.1$.

We also conduct simulation analysis to examine the impact of social influence. In our numerical analysis, 20% of prediction market participants are experts and they have more precise private signals (the precision is twice as the precision of private signals of ordinary participants). In this case, people will place larger weights on the information from these experts. The simulation results in a benchmark non-networked market and in a regular social network (k = 2) are presented in Figures D.2 and D.3, respectively.

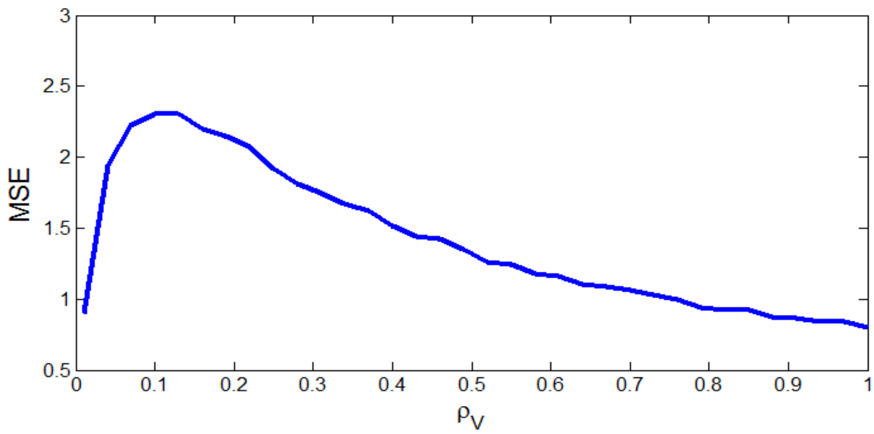


Figure D.2 The Effect of the Precision of Public Information in a Non-Networked Prediction Market: Experts vs. Ordinary Participants

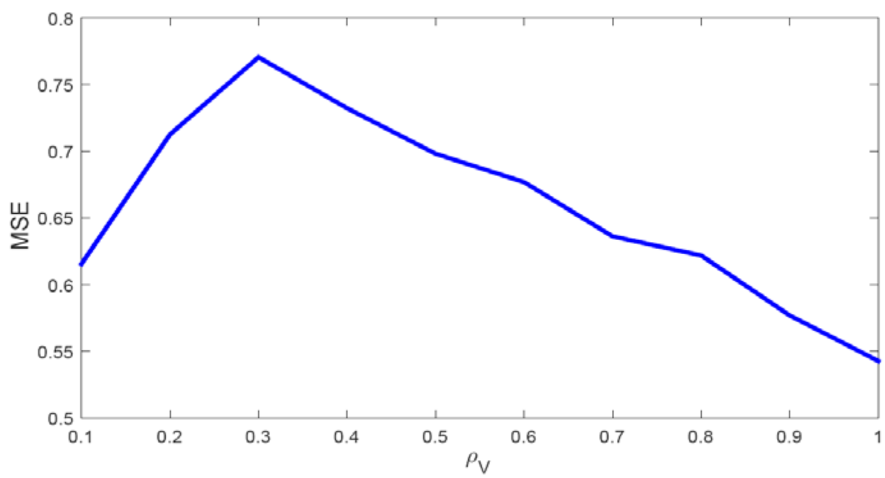
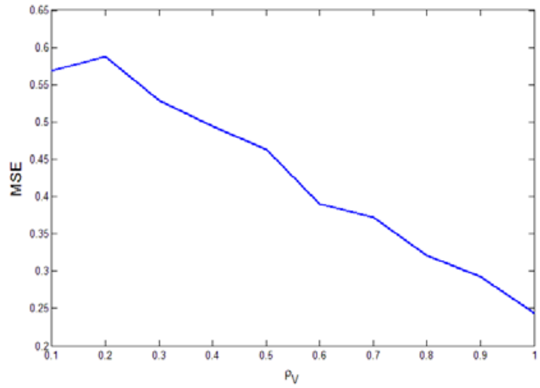
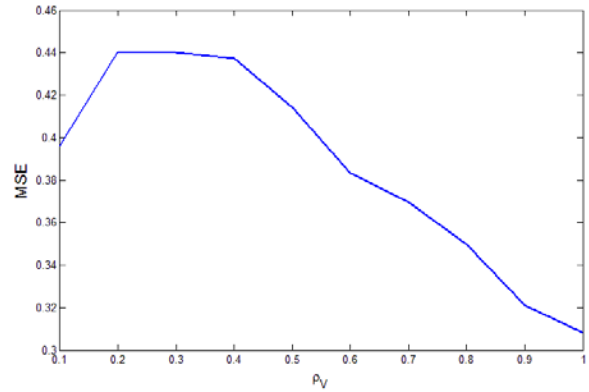


Figure D.3 The Effect of the Precision of Public Information in a Regular Social Network: Experts vs. Ordinary Participants

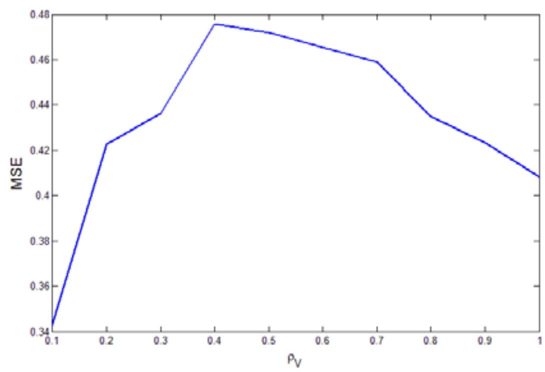
The simulation results with the heterogenous precesion setting in additional complicated social networks are presented in Figure D.4. We find that our results are robust: greater public information precision may be detrimental to prediction market accuracy.



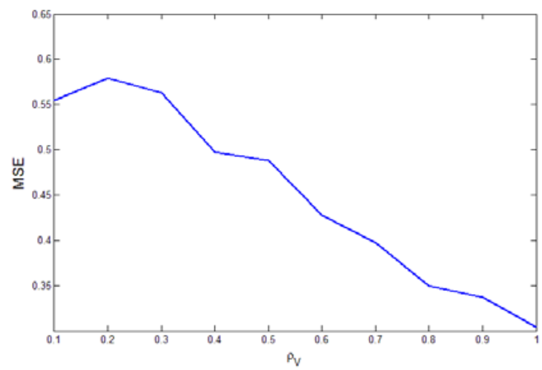
(a) Gilbert Network



(b) Erdos–Renyi Network



(c) Small-World Network



(d) Preferential Attachment Network

Figure D.4 The Effect of the Precision of Public Information in Complicated Social Networks: Experts vs. Ordinary Participants

Online Appendix E: Forecast-Report Prediction Market Mechanism

In real-world prediction markets, there are two commonly used mechanisms of information aggregation: a security-trading mechanism and a forecast-report mechanism (Jian and Sami 2012). A security-trading mechanism is similar to a competitive financial market, and people trade securities based on their forecasts. The market clears when the aggregate demand for securities equals the supply, and market clearing determines the prediction market price. In this paper, we focus mainly on the security-trading mechanism.

A forecast-report mechanism is a proper scoring rule that elicits the true beliefs of participants as probabilistic forecasts. The proper scoring rules give the participants the incentives to report truthfully, then the principal aggregates the private information of all participants. For instance, the Ford Prediction Exchange (FPEx) was the first prediction market at Ford, developed in 2006. Instead of buying and selling stock, it used a scored polling mechanism in which traders made forecasts by specifying the individual predictions (Montgomery et al. 2013). In this appendix, we show that the overweight issues still exist in a forecast-report prediction market mechanism.

The basic model setup of a forecast-report mechnaism is similar to that in Section 3.1. All the prediction market participants share a common prior on V, given by:

$$V \sim N(V_0, 1/\rho_V),$$

Before the prediction market opens, each participant can access a private signal:

$$S_i = V + \varepsilon_i, \varepsilon_i \sim N(0,1/\rho_\varepsilon), \varepsilon_i \perp \varepsilon_j.$$

The manager designs a quadratic loss function to elicit the private information of prediction market participants. A participant's payoff function is given by:

$$w(x_i, V) = a - b(x_i - V)^2,$$

where x_i is the prediction reported by participant i, and $b(x_i - V)^2$ is a quadratic penalty term for mistakes in the forecast. The optimal report for participant i is $x_i^* = E[V|I_i] = E[V|S_i]$, where I_i is the information set of participant i.

Following the prior literature (Armstrong 2001), we assume that the manager adopts a simple averaging rule to aggregate all participants' forecasts, and his prediction is:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^* = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[V|I_i] = \frac{\rho_V}{\rho_\varepsilon+\rho_V}V_0 + \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}S_i$$

$$= \frac{\rho_V}{\rho_\varepsilon+\rho_V}V_0 + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}V + \frac{\rho_\varepsilon}{\rho_\varepsilon+\rho_V}\bar{\varepsilon}.$$

The weight on public information in a forecast-report mechanism is given by:

$$W_F = \frac{\rho_V}{\rho_\varepsilon + \rho_V} \geq W_m = \frac{\rho_V}{n\rho_\varepsilon + \rho_V},$$

where W_m is the efficient weight on public information. Therefore, the issue of overweighting public information still exists in a forecast-report prediction market mechanism.

Online Appendix F: Trade-off between Information Precision and Information Diversity

As argued in Keuschnigg and Ganser (2016), crowd wisdom does not only depend on the prediction ability/precision of agents, but also depends on the information diversity. In our simulation analysis, we examine the trade-off between information precision and information diversity by looking at two departments within a company. In Department H, each employee can access a high precision signal with $\rho_{\varepsilon H} = 0.15$, while in Department L, each employee receives a low precision signal with $\rho_{\varepsilon L} = 0.1$. In other words, employees in Department H have more precise information on this specific prediction market topic. For instance, employees in the marketing department of a company may have more precise information on product sales. In order to capture correlated information sources within a department, we assume that the private signal errors of two employees in a same department are positively correlated, but are independent if they are from different departments.

We consider an optimal selection problem of prediction market participants. Suppose that a corporate manager wants to a build a prediction market with $n = 50$ participants, and all prediction market participants will be chosen from either Department H or Department L (without loss of generality, we assume that each department has 50 employees). In other words, $n_H + n_L = 50$, where n_H is the number of participants chosen from Department H, and n_L is the number of participants chosen from Department L. In the following simulation analysis, we examine the impact of information diversity on the composition of prediction market participants. For simplicity, we set parameter values $V_0 = 10$, $\rho_V = 0.1$ and $k = 1$. Since we are interested in the impact of information diversity, we vary the correlation coefficient of private signal errors of employees in a same department: $\delta = 0, 0.3, 0.6, 0.9$. Under each correlation coefficient, we run the simulation 10,000 times to compute the optimal number of participants chosen from Department H, n_H^*, that achieves the highest prediction performance (the lowest MSE), and plot the following figure.

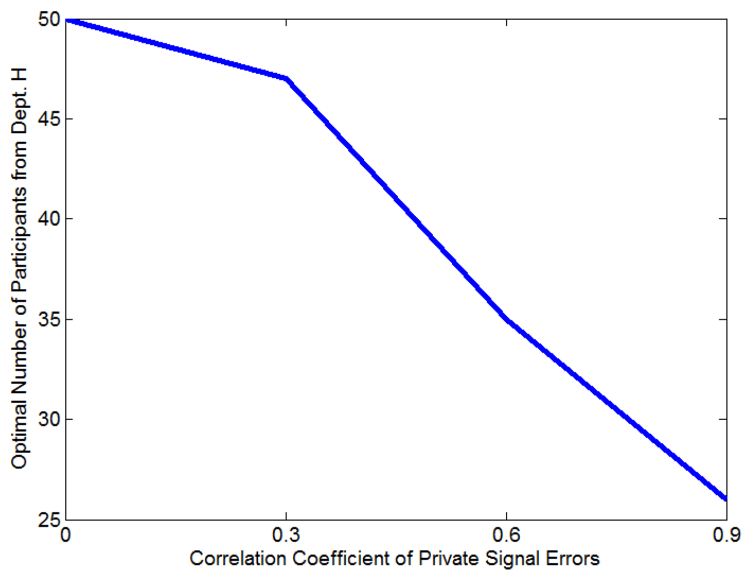


Figure F.1 The Trade-off between Information Precision and Information Diversity

Apparently, when the correlation coefficient $\delta = 0$, all prediction market participants should come from Department H. The reason is that when information within a department is not correlated, the effect of information precision dominates: The manager should choose employees with the highest prediction precision. As the correlation coefficient increases, we find that the optimal number of participants chosen from Department H, n_H^*, decreases, which shows a clear trade-off between information precision and information diversity. When δ is high, the information sources within a same department are highly correlated. Although Department *H* employees have more precise information, it is beneficial to have some Department *L* employees as diverse information sources.

References

Armstrong, J. S. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell: Kluwer Academic, 2001.

Jian, L., R. Sami. (2012). Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution. *Management Science* 58(1) 123–140.

Keuschnigg, M., Ganser, C. (2016). Crowd Wisdom Relies on Agents' Ability in Small Groups with a Voting Aggregation Rule. *Management Science*, forthcoming.

Montgomery, T. A., Stieg, P. M., Cavaretta, M. J., Moraal, P. E. (2013). Experience from hosting a corporate prediction market: benefits beyond the forecasts. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1384-1392). ACM.