RESEARCH ARTICLE

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}
and 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}

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 MSE(P^*) decreases with n. We can also obtain

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

and

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

Therefore, if $\frac{\rho_V}{\rho_\varepsilon} \le \frac{n-2}{n}$, $\frac{\partial}{\partial \rho_V}\mathrm{MSE}\left(P^*\right) \ge 0$, if $\frac{\rho_V}{\rho_\varepsilon} > \frac{n-2}{n}$, $\frac{\partial}{\partial \rho_V}\mathrm{MSE}\left(P^*\right) < 0$.

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, $\frac{\rho_V}{\rho_\varepsilon} \le \frac{n-2}{n}$, 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 coefficients *a*, *b*, and *c*.

Proof of Proposition 5

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

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 equilibrium 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 $\frac{\rho_V}{\rho_\varepsilon} \le \frac{2}{1+\delta}\left[\frac{n-4}{n}\right]$, $\frac{\partial}{\partial \rho_V}\text{MSE}(P^*)$ 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.

Appendix B

A Heterogeneous Social Network

In a heterogeneous social network, we have two types of participants: (1) participants whose degree is 0; and (2) 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 heterogeneous social network will degenerate to two special cases: a nonnetworked 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 heterogeneous 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 nonnetworked 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:

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

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 heterogeneous social network, the equilibrium prediction market price is given by*

$$\begin{aligned} P^* &= \frac{\frac{a_0 n}{\mathbf{Var}_0[V|I_i]}}{\frac{a_0 n}{\mathbf{Var}_0[V|I_i]}+\frac{(1-a_0)n}{\mathbf{Var}_k[V|I_j]}}\sum_{i\in D_0}\frac{\mathbf{E}_0[V|I_i]}{a_0 n} + \frac{\frac{(1-a_0)n}{\mathbf{Var}_k[V|I_j]}}{\frac{a_0 n}{\mathbf{Var}_0[V|I_i]}+\frac{(1-a_0)n}{\mathbf{Var}_k[V|I_j]}}\sum_{j\in D_k}\frac{\mathbf{E}_k[V|I_j]}{(1-a_0)n} \\ &= \delta V_0 + (1-\delta)V + \frac{\rho_\varepsilon}{[a_0+(1-a_0)(k+1)]\rho_\varepsilon+\rho_V}\left[\sum_{i\in D_0}\frac{\varepsilon_i}{n} + (k+1)\sum_{j\in D_k}\frac{\varepsilon_j}{n}\right] \end{aligned}$$

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 heterogeneous 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 nonnetworked 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 heterogeneous 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 heterogeneous social network:

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

Note that when $a_0 = 1$, the MSE in a prediction market with a heterogeneous social network will be degenerated to equation 7, the MSE in a nonnetworked 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 $\frac{\rho_V}{[[a_0+(1-a_0)(k+1)]\rho_\varepsilon+\rho_V]^2}$. 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 nonnetworked prediction market is greater than the MSE in a prediction market with a heterogeneous social network, and the MSE in a prediction market with a heterogeneous 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 heterogeneous 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 heterogeneous 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 heterogeneous 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 nonnetworked 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 heterogeneous social network. Figure B.1 depicts the contour lines of the MSE in a heterogeneous social network when $n = 50$, $k = 9$, and $a_0 = 0.8$. The marginal line in a heterogeneous social network (the solid line) is between the marginal line in a nonnetworked 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 heterogeneous network is larger than that in a nonnetworked prediction market, but smaller than that in a regular network under the chosen parameter values. The intuition is that a heterogeneous network is a linear combination of a regular network and a nonnetworked environment.

Therefore, the marginal line in a heterogeneous 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 $a_0 = 0$, $\frac{\rho_V}{\rho_\varepsilon} = (k+1)\left[\frac{n-2(k+1)}{n}\right]$.

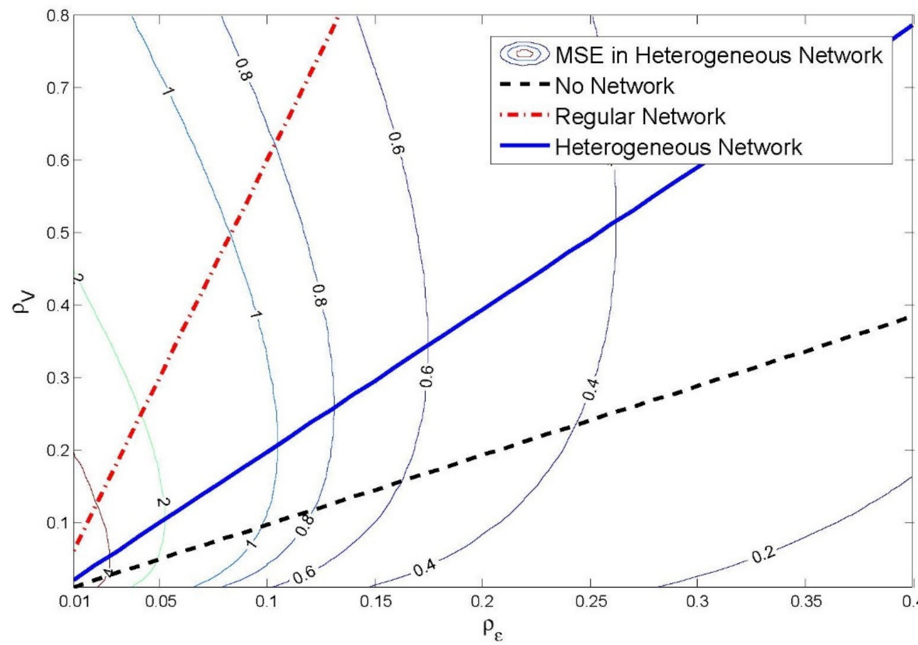

Figure B1. The Impact of Public Information Precision on Prediction Market Performance (Heterogeneous Network), *n* = 50, *k* = 9, *a*₀ = 0.8

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 nonnetworked 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 nonnetworked 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 $k = 5$: $\rho_\varepsilon = 0.025$.

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.

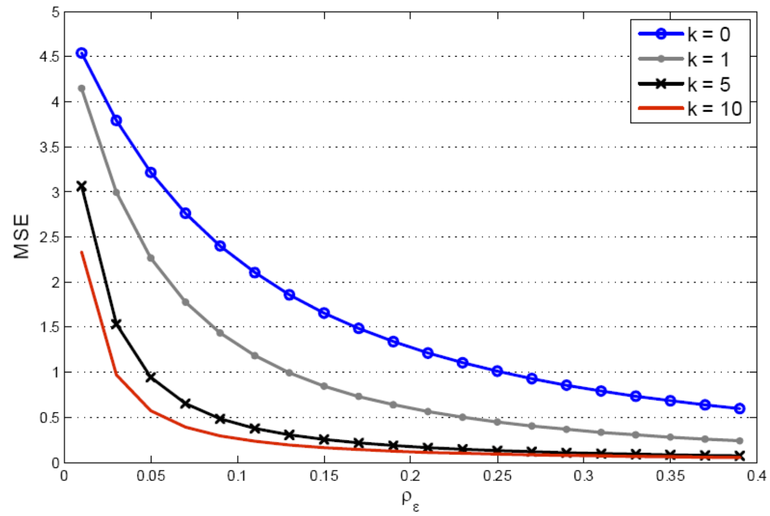

Figure C1. The Impact of Social Interaction on Prediction Market Performance, *n* = 50 and $\rho_V = 0.2$

Appendix D

Additional Numerical Analysis

To provide a benchmark, in Figure D.1, we depict the nonnetworked case using the same parameter values as those in Section 4.3. The pattern is similar, but we have two additional observations: (i) The MSE in a nonnetworked 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 nonnetworked prediction market than in a socially embedded prediction market under the chosen parameter values. This is reminiscent of Proposition 11.

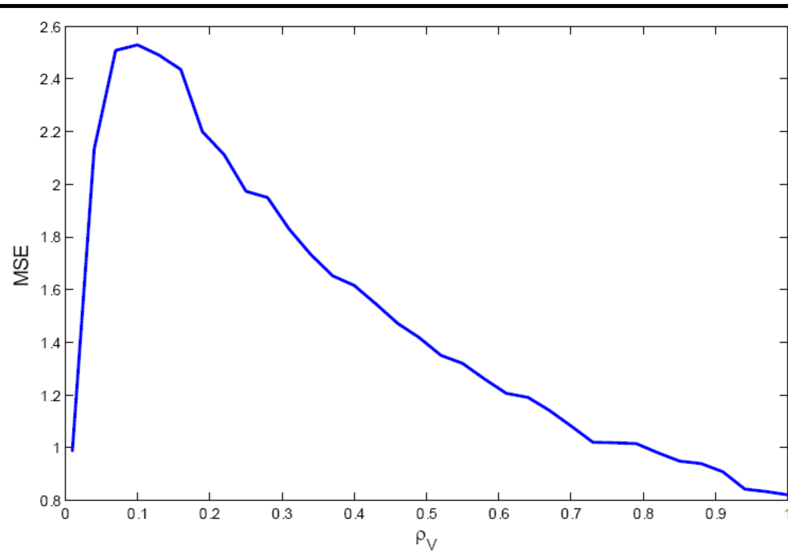

Figure D1. The Effect of the Precision of Public Information on the MSE in a Nonnetworked Prediction Market, $n = 50$, $V_0 = 10$, and $\rho_\varepsilon = 0.1$

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 nonnetworked market and in a regular social network (k = 2) are presented in Figures D2 and D3, respectively.

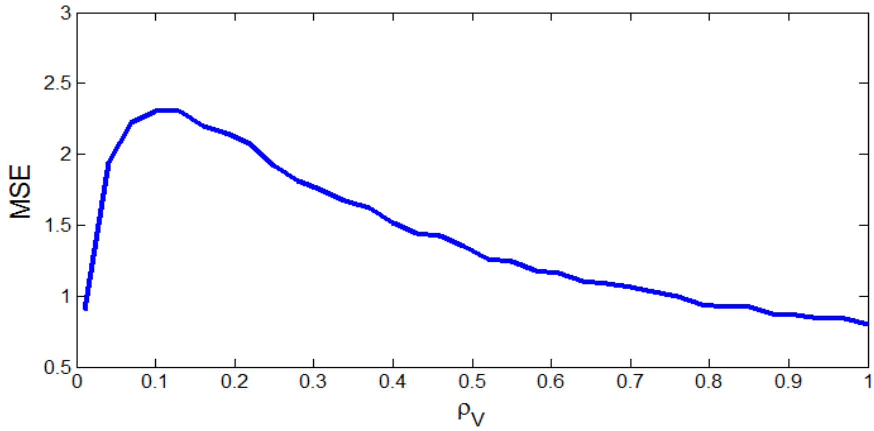

Figure D2. The Effect of the Precision of Public Information in a Nonnetworked Prediction Market: Experts Versus Ordinary Participants

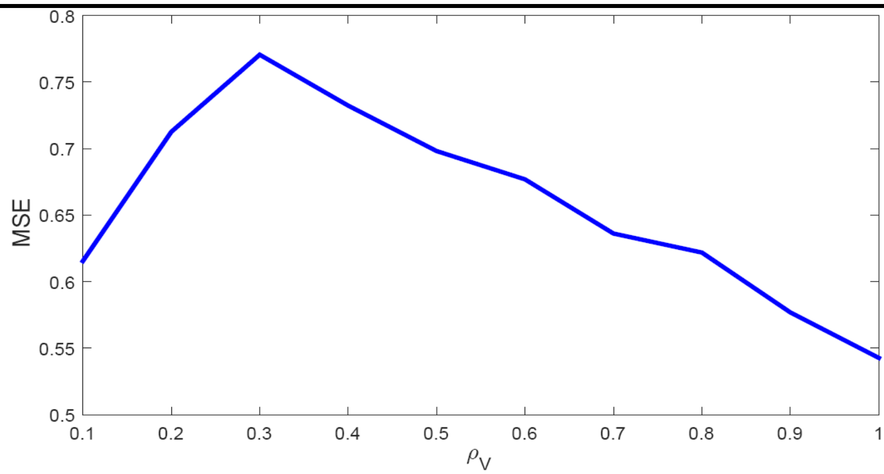

Figure D3. The Effect of the Precision of Public Information in a Regular Social Network: Experts Versus Ordinary Participants

The simulation results with the heterogeneous precision setting in additional complicated social networks are presented in Figure D4. We find that our results are robust: greater public information precision may be detrimental to prediction market accuracy.

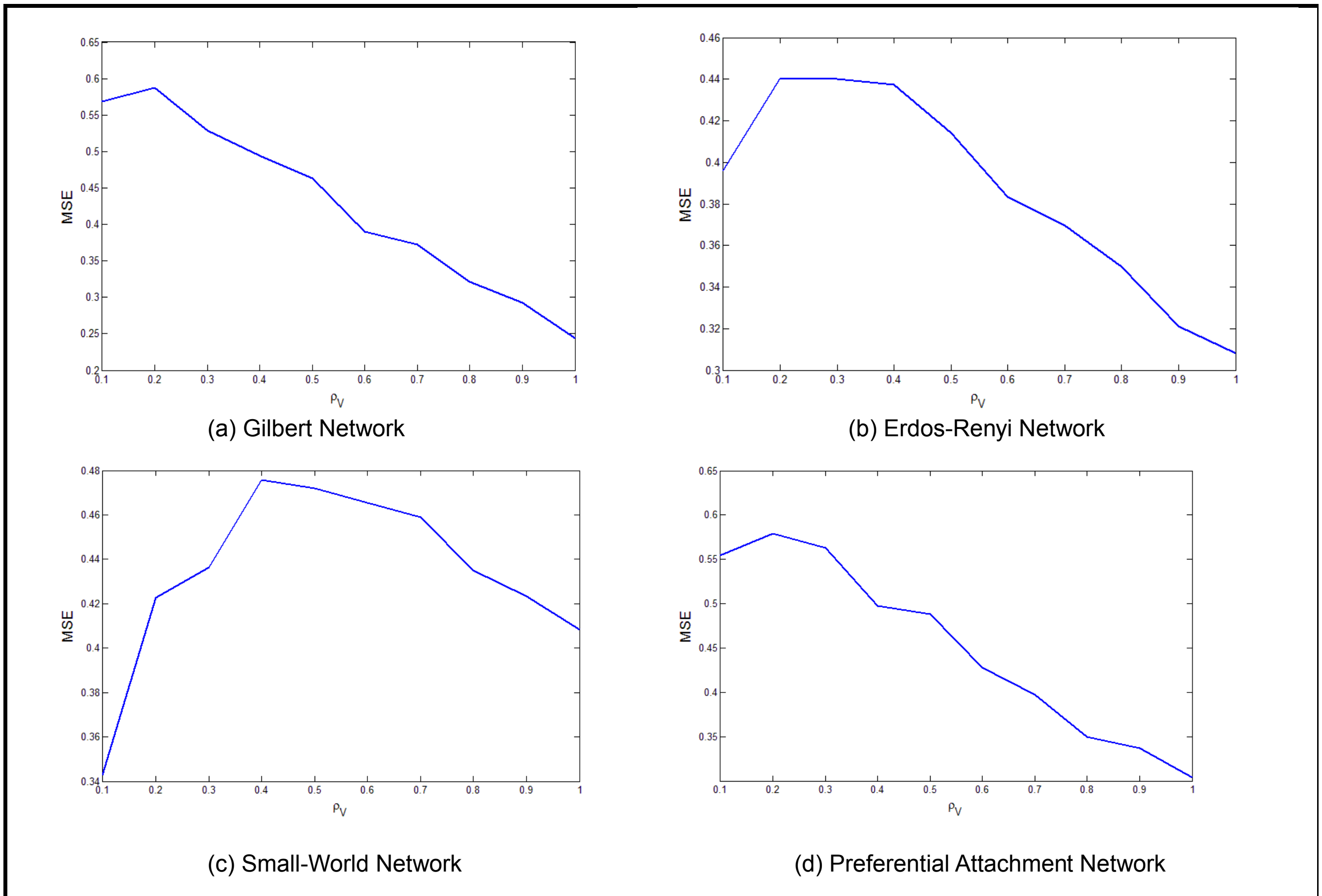

Figure D4. The Effect of the Precision of Public Information in Complicated Social Networks: Experts Versus Ordinary Participants

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 mechanism is similar to that in the “Model Setup” section. 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

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

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.

Appendix F

Trade-Off between Information Precision and Information Diversity

As argued in Keuschnigg and Ganser (2017), 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.

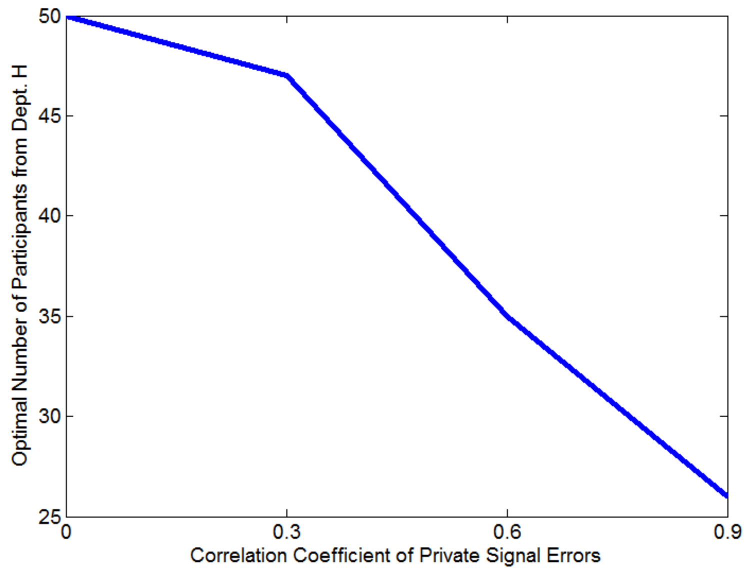

Figure F1. The Trade-Off between Information Precision and Information Diversity

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

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

Keuschnigg, M., and Ganser, C. 2017. "Crowd Wisdom Relies on Agents' Ability in Small Groups with a Voting Aggregation Rule," *Management Science* (63:3), pp. 818-828.

Montgomery, T. A., Stieg, P. M., Cavaretta, M. J., and 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*, New York: ACM Press, pp. 1384-1392.