Three Essays on Professional Advising

by

Kyoung-Soo Yoon

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2008

Doctoral Committee:

Professor Lones A. Smith, Co-chair
Associate Professor Uday Rajan, Co-chair
Associate Professor Emre Ozdenoren
Assistant Professor Stephen Lauerman
Assistant Professor Xuenan Li

To my wife, Kay.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to my advisors, Professor Lones Smith and Professor Uday Rajan. Professor Lones Smith has taught me what the theorist would be and how wonderful it is to research as a theorist. He has guided me with his great intuition whenever my research came to a deadlock. Professor Uday Rajan has taught the way a theorist turns real economic problems into theoretical models. His broad knowledge on financial issues and logical way of thinking have been of great value for me. Their understanding, encouragement, and personal guidance have provided a solid foundation for this dissertation.

I am deeply grateful to my committee members; Professor Emre Ozdenoren, Professor Stephan Lauerman, and Professor Xuenan Erica Li for detailed and constructive comments, helpful suggestions, and encouragements throughout this work.

Chapter 1 and chapter 2 of the dissertation are joint work with Jooyong Jun. His profound knowledge of methodology and his devotion to research were instrumental in those chapters.

Throughout my seven years, I was supported for many semesters by the Department of Economics at the University of Michigan working as a graduate student instructor.

Finally, I would like to dedicate this dissertation to my wife, Kay, who has been with

me through the journey toward my doctoral degree, and has been a mentor in my life.

PREFACE

Information is often unequally dispersed among people, and people seek advice from experts. This dissertation considers three topics related to professional advising and forecasting.

Chapter 1 and 2 are joint essays with Jooyong Jun. We consider an uninformed principal wishing to give the best prediction of the true state by hiring informed experts. The precision of the private signal of each expert is heterogeneous and unknown to the principal. The principal should decide the compensation scheme and employment policy. In this situation, screening the precision of each expert is valuable since it allows the principal to aggregate information efficiently and to make a clear cut employment policy.

Chapter 1 presents a model in which a principal tries to design the compensation scheme optimally to screen the type of each expert. Under a Gaussian specification, it is shown that there exists a payoff function which achieves the first best outcome: each expert is induced to report honestly on the true state, truthfully revealing his own type, and is paid only his reservation utility. The optimal contract is a linear function with respect to the mean squared error to the power of a certain degree. The result comes from 1) the single crossing property of the linear payoff function, and 2) the cheap talk feature of the professional advising.

Chapter 2 explores the optimal employment with experts. Assuming Gaussian noise, the objective function of the principal becomes a function of the sum of types of employees and is submodular and monotone in the sets of experts. Thus, the production exhibits decreasing marginal returns. We transform the production function from a set function to a function on Euclidian space, depending on the current employment set and the type of an additional employee. The marginal production function decreases more quickly for higher types than lower ones, as the current employment set is enlarged.

The property of the optimal employment set depends on the reservation utility. It is shown that if the reservation utility is proportional to the marginal single production, the optimal employment set follows a cut-off property in which experts with higher precision are employed. In this case, sequential hiring from the highest type leads to the global optimum.

The chapter also develops a few extensions of the main result. We show that the cut-off property holds with a general set production function when it is submodular exhibiting the decreasing curvature property, and the reservation utility is proportional to the marginal single production. We also propose an efficient algorithm to find the optimal set with a general submodular set function.

Chapter 3 considers forecasting behavior by an expert when the arrival timing of a new signal is uncertain. The forecaster needs to infer the arrival time since he does not know whether the information he possesses is new or already reflected in the predecessors' forecasts. We analyze the Bayesian updating procedure in this situation.

Assuming Gaussian noise, it is shown that the optimal forecast depends on the deviation of the signal from the consensus. This is because, as the signal of the forecaster moves out from the consensus forecast, the forecaster assigns more weight to his own signal, believing that it is more likely to be new. This leads to a tendency towards more extreme forecasts.

The result sheds light on recent empirical studies on a herding or anti-herding bias. Without resorting to behavioral assumptions or unusual payoff functions, our model shows that statisticians may observe forecasters placing more weight on private information rather than the consensus.

Table of Contents

List of Figures

Figure

Chapter 1

Getting Advice from Experts: Optimal Contracts

1.1 Introduction

This chapter[1] explores mechanisms through which a principal can best elicit information from multiple experts. We in particular focus on a contractual situation, implicitly assuming the information the principal needs to gather is at least partially specific to him. Two important issues emerge as the principal makes contracts with multiple experts. Firstly, the principal should consider how to aggregate information from multiple sources. Secondly, she should determine the wage offer to each expert, which may depend on the information quality if available.

Efficient information aggregation crucially depends on whether the principal can screen the precision of information each expert possesses. To see this, consider an environment in which experts have heterogeneous private signal distributions and the signal is independently informative of the true state. The heterogeneity reflects different abilities in processing raw information, analytic technologies, and/or levels of 'animal spirits.'

Provided the principal knows the precision of each expert, she can easily aggregate information from multiple experts by Bayesian updating. Assuming the compensation scheme is designed so that experts are paid off according to the ex-post accuracy, and experts have no payoff other than the wage paid by the principal, each expert must submit the report on the true state at his posterior mean given his available informa-

[1]This chapter represents jointly work with Jooyong Jun.

tion[2]. Without message pooling, the one-to-one relationship between posterior mean and private signal allows the principal to discern each expert's private information. It follows that the principal's best prediction on the true state is then the weighted average of the prior and all private signals, where the weights are determined by the precision of the signal.

When the precision of each expert's signal is unknown to the principal, however, information aggregation cannot be achieved with the simple compensation scheme described above. A posterior mean is no longer matched to the private signal in a one-to-one relationship, and the principal would not know the weight she should assign to the report of each expert. The compensation must be more sophisticated. It should be designed not only to induce the honest report on the true state but also to elicit the precision of expert's report.

Sorting through compensation helps the principal via another channel. When the information quality of each expert is heterogeneous and the reservation wage depends on the quality, the principal also needs to decide the wage offer and it would be beneficial if the wage offer can be contingent on the information quality.

We show, in an environment where the reservation wage is type dependent, that there exist payoff function(s) in which the true type revelation is implemented and the honest report on the true state is induced. In addition, the compensation scheme induces the first-best outcome in the sense that no information rent exceeding the reservation utility is paid in equilibrium. When the reservation wage is convex in type, a simple linear payoff function with respect to the mean squared error of the report on true state achieves the first-best. In the case when the reservation utility is concave, the optimal payoff function is more complicated but keeps the linearity in a certain form of performance measure.

[2]Experts' payoff other than the wage from the principal, including nonpecuniary or implicit compensation, may drive shaded or pooled messages on true state. For example, reputation concern induces experts to shade or pool messages in the model of Ottaviani and Sørenson (2006a, 2006b).

The intuition behind this sorting mechanism is straightforward. In the optimal compensation scheme we propose, the principal asks each expert what his type is. The optimal contract is designed so that the penalty for the incorrect report is increasing in type announced. The less accurate expert then incurs more cost when he pretends to be more accurate type, barring untruthful type revelation. Moreover, due to the cheap talk feature of the 'production' of advice, there is no intrinsic utility or cost for experts. This implies the virtual surplus is linear in the control variable of the principal, and the principal makes the information rent arbitrarily small up to the reservation utility to achieve sorting.

We then propose a game in which the principal achieves not only the efficient information aggregation but also the optimal employment. In the game, the principal announces the payoff function which depends on the type, the precision of the private signal announced by each expert, the report on true state, and the true state to be revealed ex-post. Experts from population then apply for the job (pre-screening stage.) Among those applicants, the principal decides which experts to hire (employment stage.) Compensation is paid after the true state is revealed.

This chapter is organized as follows: A brief literature survey follows the introduction. We describe the model and present the optimization problem in the next section. In section 3, we derive the optimal contract in which honest reporting and truthful type revelation are achieved and the participation constraint is binding. In section 4, we propose a game to achieve the optimal employment. Section 5 concludes and addresses issues for further research.

1.2 Related Literature

The sorting mechanism in the paper is an application of a screening problem under asymmetric information. For example, Maskin and Riley (1984) address the problem in the context of an optimal quantity discount by a monopolist. The main difference is that in professional advising, the information asymmetry occurs not only in the type of each agent but also in the true state which is realized ex-post. Indeed, the type, or the information quality of each expert is revealed ex-post through the realized true state and the forecast. The principal, therefore, needs to get messages from each agent about her type in addition to the forecast on the true state. In a sense, the model presented here is a hybrid model of screening and moral hazard because the latter message is often sent after the principal's employment decision is made.

Bhattacharya and Pfleiderer (1985) is more directly related to our work in the motivation and the model specification. They examine the compensation problem for risk-averse portfolio managers whose signal and signal distribution are both private information. They also derive the compensation scheme which achieves the first best outcome. It differs from ours in the objective function and the risk attitude. They assume the utility function of both principal and agents exhibits constant absolute risk aversion, which makes sense in the context of the delegation of portfolio management. With risk neutral agents, as in our model, the problem is not well defined since the portfolio choice position would be extreme. In this sense, the first main result of this paper is a risk neutral agent version of section 4 in Bhattacharya and Pfleiderer (1985). The second main theorem is new. While Bhattacharya and Pfleiderer (1985) derives the first best outcome under some regularity conditions on the reservation utility, we show it for quite general case by varying the performance measure.[3]

Crémer and McLean (1985, 1988) study mechanisms in which a principal, or a seller, extracts full surplus in the context of the independent value auction. In their model, the valuations of the bidders are correlated and they know this fact. The seller then designs an auction mechanism in which payments depend on the types announced by bidders. Under some regularity conditions, the seller can induce each

[3]Osband (1989) also studies the incentive provision problem for forecasters. The precision of each forecaster in his model depends on the effort level, so the focus is on the moral hazard problem, not on the hidden type problem as ours.

bidder to announce his type truthfully, which results in the full surplus extraction.

The types of experts in our model are also correlated, but they are *conditionally* independent. The true state itself, which is assumed to be verifiable ex-post, becomes a reference point that each expert's type is measured. Each expert, thus, is induced to announce his type truthfully without guessing other experts' type. This allows us to develop an *independent* compensation scheme that does not depend on the type announcements by other experts. Auctions with state-dependent payments are studied in Hansen (1985), but it deals with very special cases.

Recent literature on professional advisors is based on the cheap talk game model first introduced by Crawford and Sobel (1982). Departing from partisan bias exogenously given to the payoff functions, Scharfstein and Stein (1990) explores how reputation concerns affect the pattern of messages in equilibrium. They show the reputation concern drives experts to herd in a binary model. The model is generalized in Ottaviani and Sørensen (2006a, 2006b).

The main difference between this paper and the previous literature on professional advising is twofold. First, we give the principal an active role in determining the compensation scheme. Secondly, our focus is on efficiency in information aggregation and employment, not on the strategic bias. To do so, we assume the principal has no private information, and experts are not concerned the reputation effect of the current report.

In our model, the information asymmetry is two dimensional: the signal and its distribution. Except for a few papers, most existing papers on professional advising assume experts share the common private signal distribution, and asymmetric information lies only in realized value of their private signal. Avery and Chevalier (1999), Levy (2004), and Ottaviani and Sørensen (2006a) consider heterogeneous private signal distribution but usually the uncertainty is assumed to be symmetric across the players in the model. Trueman (1994) and section 6 in Ottaviani and Sørensen

(2006a) model asymmetric information on signal distribution. The information structure in this paper is mostly similar to Ottaviani and Sørensen (2006a). Battaglini (2002) explores a cheap talk game with multi dimensional uncertainty and multiple referrals, but his results are mainly derived from the orthogonality between uncertain variables, which is different from our setting.

1.3 Model

An uninformed principal tries to make his best prediction of the true state, for example the profitability of a project. To get better information, the principal wishes to hire privately informed agents, who are called 'experts' hereafter. Experts are heterogeneous in the precision of their private signal, which is labeled their *type*. The principal designs a game as follows.

The true state x is drawn from a normal distribution with mean μ_x and variance $1/\tau_x$, which is common knowledge. We assume the true state is verifiable and thus contractible. While the principal has no private information[4], expert $i \in I$ receives a conditionally independent private signal s_i, where I is the set of experts who are employed.[5] The distribution of s_i is assumed to follow a normal distribution with mean x and precision τ_i, or

$$s_i = x + \epsilon_i, \quad \epsilon_i \sim N\left(0, 1/\tau_i\right).$$

The precision, or type, of each expert τ_i is drawn from the population with distribution function F on the support of $[\underline{\tau}, \overline{\tau}] \subset \mathbb{R}^+$. We assume F is continuously differentiable so that the probability continuous density function f exists. The principal cannot discern the type of each agent, but each expert knows his own type. In

[4]The assumption of a fully uninformed principal, in addition to that of the payoff being conditioned on the true state, precludes the 'yesman effect' in Prendergast (1993).

[5]We fix the employment set of experts in this chapter, as though the employment decision is made before the contract and the information aggregation. However, the order may be reversed in order for the contract to be used as a pre-screening device. The whole recruiting, contracting, and information aggregation process is discussed in the later section.

the pre-screening stage, each expert is requested to submit a message on his own type t_i. Once hired, he has to submit a *report* on the true state, denoted by $r_i \in \mathbb{R}$ for expert i.[6]

The principal's objective is to maximize *revenue* less payoffs to employed experts. The revenue function R depends on the principal's prediction on the true state, denoted by $\widehat{x}$, and the true state. We assume the revenue is decreasing in the ex-post error, $\mid \widehat{x} - x \mid$. For example, the revenue can be the negative mean-squared error where $R(x, \widehat{x}) \equiv -a(\widehat{x} - x)^2$ for a constant $a > 0$. In this case, the revenue function is a decreasing function of the mean squared error.

The only cost for the principal is the wage she pays to the experts, where the payoff function is denoted by $C(r_i, t_i, x)$. Note that the payoff does not depend on other experts' messages. In other words, we restrict the compensation to be independent, which implies that the principal cannot use group incentives to implement the information revelation and the performance must be evaluated through the absolute performance basis[7].

Experts are assumed to be risk neutral utility maximizers with the identical vNM utility function $u(c) = c$. We assume that the only benefit from information provision is the payoff from the principal. Each expert has a reservation utility which is type-dependent. Type-dependent reservation utility function $\underline{u}(\tau)$ is assumed to be strictly

[6]We follow the convention that each expert reports his best prediction, not directly revealing his private signal. However, reporting the prediction is equivalent to reporting the signal in equilibrium provided there is no message pooling, which is the case of this paper.

[7]Relative performance evaluation has been an important issue in contract theory with multi-agent models. We exclude such evaluation on report for simplicity and tractability of the payoff function, since we are focusing on the screening procedure. Extant literature in contract theory find the merits of relative performance evaluation in that it reduces risk-sharing cost as to the common noise. See Holmstrom (1982). Another branch of literature regarding relative evaluation studies rank order compensation or tournament. While it has been shown that tournament scheme can provide approximately the same incentive for agents as the standard contractual form(See, for example, Green and Stokey (1983)), it is less susceptible to extreme output volatilities. Both benefits mentioned above are not relevant to the current model. Ottaviani and Sørenson (2006b) consider forecasting contest, an extreme case of compensation scheme based on relative performance evaluation, in the context of reputational cheap talk game, but their information structure is different from ours.

increasing and continuously differentiable. Increasing reservation utility is realistic when the private information is not fully relation-specific. The expert might use the private information outside of the principal-agent relationship to derive some personal benefits from it.[8]

The Principal's action is denoted by $(\widehat{x}, C)$. The optimization problem is formally described as follows.

$$\max_{\widehat{x},C} E_x \left[R(\widehat{x}, x) - \sum_{i \in I} C(r_i, t_i, x) \right]$$

subject to the expert's problem

$$(r_i, t_i) \in \arg\max_{r,t} E_x \left[C(r, t, x) \mid s_i, \tau_i \right]$$

subject to the participation constraint

$$E_x \left[C(r_i, t_i, x) \mid s_i, \tau_i \right] \geq \underline{u}(\tau_i).$$

In the next section, we begin our analysis with the pre-screening stage.

1.4 Compensation Scheme for Sorting

In this section, we aim at finding a compensation scheme which achieves the first-best. We ask whether there exists a payoff function $C(r_i, t_i, x)$ which induces the expert to report his posterior mean, implements him to message his own type, and further the expected payoff is just his reservation utility[9]. Formally, we want to find C satisfying

$$\left(\frac{s_i \tau_i + \mu_x \tau_x}{\tau_i + \tau_x}, \tau_i \right) \in \arg\max_{t,r} E_x \left[C(r, t, x) \mid s_i, \tau_i \right]$$

and

[8]See Jullien(2006) for examples of type-dependent reservation utility and the general solution in the context of screening problem. Bhattacharya and Pfleiderer (1985) also assume the type dependent reservation utility.

[9]We assume the massage related to the true state is the posterior mean of each expert. However, it is equivalent to assume that the private signal itself is reported.

$$E_x\left[C\left(\frac{s_i\tau_i+\mu_x\tau_x}{\tau_i+\tau_x},\tau_i,x\right)\mid s_i,\tau_i\right]=\underline{u}(\tau_i).$$

Note that once the sorting and truthful reporting are implemented, the principal's optimal action is straightforward. Provided the revenue $R(\widehat{x},x)$ is decreasing in error $\mid \widehat{x}-x\mid$, for a fixed I, the best prediction on the true state, $\widehat{x}^*$, is the posterior mean given reports from $|I|$ experts, where $|A|$ is the number of elements in a set A. Formally, we have

$$\widehat{x}^*=\frac{\sum_{i\in I}r_i(\tau_i+\tau_x)-(|I|-1)\mu_x\tau_x}{\sum_{i\in I}\tau_i+\tau_x}.$$

For example, if the revenue is negative mean squared error, the resulting expected net gain is

$$E\left[R(\widehat{x}^*,x)\right]-\sum_{i\in I}\underline{u}(\tau_i)=-\frac{a}{\tau_x+\sum_{j\in J}\tau_j}-\sum_{j\in J}\underline{u}(\tau_j).$$

Our strategy to show existence is as follows. We first restrict to a subclass of payoff functions. We then solve the standard screening problem within the class and check whether the participation constraint is binding for all types of expert.

Proposition 1 is our first main result. It states that if the reservation utility function is non-convex, the first-best outcome is achieved through a payoff function which is linear in the mean squared error of the report. With the linearity restriction, the expert with precision τ_i should solve

$$(r_i,t_i)\in\arg\max_{t,r}E_x\left[-\alpha(t)(r-x)^2+\beta(t)\mid s_i,\tau_i\right] \tag{1.1}$$

We want to find $\alpha(t)$ and $\beta(t)$ such that the solution to (1.1) satisfies the conditions for both honest reporting and truthful type revealing, as well as the participation constraint. We must therefore have the following conditions:

- Incentive Compatibility for Honest Reporting (ICR)

$$\alpha(t)\geq 0. \tag{1.2}$$

- Incentive Compatibility for Truthful Type Revelation (ICT)

  Given (1.2), the expert with precision τ_i will solve the following problem:

$$\tau_i \in \arg\max_{t_i} -\frac{\alpha(t_i)}{\tau_x + \tau_i} + \beta(t_i). \tag{1.3}$$

- Participation Constraint (PC)

  Once (1.2) and (1.3) are satisfied, the participation constraint for type τ_i becomes

$$\max_{r_i, t_i} E_x \left[-\alpha(t_i)(r_i - x)^2 + \beta(t_i) \right] = -\frac{\alpha(\tau_i)}{\tau_x + \tau_i} + \beta(\tau_i) \geq \underline{u}(\tau_i). \tag{1.4}$$

PROPOSITION 1.1: *Suppose the reservation utility is convex on the support of τ. Then, the first-best is strictly implemented through the payoff function within the class of linear functions in mean squared error. That is, it is the strict best response for each expert to message his own type and submit his posterior mean, and the payoff is only his reservation utility if the payoff function is designed to be*

$$C(r, t, x) = -\alpha(t)(r - x)^2 + \beta(t)$$

where

$$\alpha(t) = (\tau_x + t)^2 \underline{u}'(t)$$

and

$$\beta(t) = (\tau_x + t)\underline{u}'(t) + \underline{u}(t).$$

PROOF. Let α and β be $\mathcal{C}^2$ function on $\mathbb{R}_{++}$.[10] Let $\alpha(t) \geq 0$ to satisfy (ICR). Define $\mathbf{C}(\tau, t)$ be the expected payoff when type τ expert announces that his type is t and

[10]It is required that $\alpha'(t)$ and $\beta'(t)$ are right continuous at 0.

he reports posterior mean honestly. Given (ICR), we have

$$\begin{aligned}\mathbf{C}(\tau,t) &= E_{x,s}\left[C(\frac{\tau s+\tau_x\mu_x}{\tau+\tau_x},t,x)\mid\tau,s\right]\\ &= -\alpha(t)E_{x,s}\left(\frac{\tau s+\tau_x\mu_x}{\tau+\tau_x}-x\right)^2+\beta(t)\\ &= \frac{-\alpha(t)}{\tau+\tau_x}+\beta(t)\end{aligned}$$

The first order condition for (ICT) is then

$$\forall\tau\in[\underline{\tau},\overline{\tau}],\quad -\frac{\alpha'(\tau)}{\tau_x+\tau}+\beta'(\tau)=0. \tag{1.5}$$

To see the second order condition given (IRC) and the first order condition of (ICT), consider the following formula.

$$\frac{\partial\mathbf{C}}{\partial t}(\tau,t)=-\alpha'(t)\cdot\frac{t-\tau}{(\tau_x+\tau)(\tau_x+t)}, \tag{1.6}$$

which implies that $t=\tau$ is the global maximizer of $\mathfrak{C}(\tau,t)$ for all τ if and only if $\alpha(t)$ is nondecreasing. We will temporarily ignore (1.6) to solve for the optimal contract with (1.5), and then check whether the contract satisfies (1.6).

Let $c(\tau)$ be the utility of expert type τ at the optimum, so that $c(\tau)=\mathbf{C}(\tau,\tau)=-\alpha(\tau)/(\tau_x+\tau)+\beta(\tau)$. Note that from envelope theorem,

$$c'(\tau)=\frac{\alpha(\tau)}{(\tau_x+\tau)^2}\geq 0 \tag{1.7}$$

which implies that the experts with higher precision are paid more.

From (1.5), we have

$$\begin{aligned}\beta'(\tau) &= \frac{\alpha'(\tau)}{\tau_x+\tau}\\ &\Rightarrow\\ \beta(\tau) &= \beta(\underline{\tau})+\int_{\underline{\tau}}^{\tau}\frac{\alpha'(s)}{\tau_x+s}ds\\ &= \beta(\underline{\tau})-\frac{\alpha(\underline{\tau})}{\tau_x+\underline{\tau}}+\frac{\alpha(\tau)}{\tau_x+\tau}+\int_{\underline{\tau}}^{\tau}\frac{\alpha(s)}{(\tau_x+s)^2}ds\end{aligned} \tag{1.8}$$

and

$$c(\tau) = -\frac{\alpha(\tau)}{\tau_x + \tau} + \beta(\tau) = \beta(\underline{\tau}) - \frac{\alpha(\underline{\tau})}{\tau_x + \underline{\tau}} + \int_{\underline{\tau}}^{\tau} \frac{\alpha(s)}{(\tau_x + s)^2} ds.$$

Given that the expert reports his posterior mean, the principal's problem is

$$\begin{aligned}
\min_{\alpha(\tau)} E_\tau \left[c(\tau)\right] &= \int_{\underline{\tau}}^{\overline{\tau}} \left(\beta(\underline{\tau}) - \frac{\alpha(\underline{\tau})}{\tau_x + \underline{\tau}} + \int_{\underline{\tau}}^{\tau} \frac{\alpha(s)}{(\tau_x + s)^2} ds \right) f(\tau) d\tau \\
&= \int_{\underline{\tau}}^{\overline{\tau}} \left(\beta(\underline{\tau}) - \frac{\alpha(\underline{\tau})}{\tau_x + \underline{\tau}} + \frac{\alpha(\tau)}{(\tau_x + \tau)^2} \frac{1 - F(\tau)}{f(\tau)} \right) f(\tau) d\tau \qquad (1.9)
\end{aligned}$$

subject to the participation constraint

$$c(\tau) = \beta(\underline{\tau}) - \frac{\alpha(\underline{\tau})}{\tau_x + \underline{\tau}} + \int_{\underline{\tau}}^{\tau} \frac{\alpha(s)}{(\tau_x + s)^2} ds \geq \underline{u}(\tau). \qquad (1.10)$$

We can solve this problem through point-wise minimization. Since the formula in the bracket of (1.9), the so called *virtual cost*, is linear in α, the principal can let $\alpha(\tau)$ be the least possible cost $c(\tau)$ for all τ. That is, the participation constraint (1.10) should bind for all τ. Differentiation of the binding participation constraint gives

$$\alpha(t) = (\tau_x + t)^2 \underline{u}'(t)$$

and from (1.8),

$$\beta(t) = \beta(\underline{\tau}) - \frac{\alpha(\underline{\tau})}{\tau_x + \underline{\tau}} + (\tau_x + t)u'(t) + \underline{u}(t) - \underline{u}(\underline{\tau}).$$

Let $\beta(\underline{\tau}) = (\tau_x + \underline{\tau})u'(\underline{\tau}) + u(\underline{\tau})$ for the participation constraint of the lowest type to bind. We now solve for α and β to satisfy the first order condition of (ICT) and binding participation constraint (PC). Finally we need to check the second order condition, which is equivalent to α monotone nondecreasing. Since

$$\alpha'(t) = 2(\tau_x + t)u'(t) + (\tau_x + t)^2 u''(t),$$

the second order condition is satisfied provided u is non-concave. This completes the proof.[11] ■

[11] The result still holds in the case of type-independent reservation utility. Suppose w is the

The following examples show how the concave reservation utility function obstructs the truthful type revelation.

EXAMPLE 1.1: *Suppose* $\tau_x = 1$ *and* $\underline{u}(\tau) = \tau$ *for* $\tau \geq 0$. *From Proposition-1, we have* $C(r,t,x) = -(1+t)^2(r-x)^2 + (1+2t)$, *and since (ICR) is satisfied,* $E_x[C(r,t,x)] = -(1+t)^2/(1+\tau) + (1+t)$. *The first order condition gives* $t = \tau$ *and the second order condition is satisfied.*

Suppose now $\tau_x = 1$ *and* $\underline{u}(\tau) = 1 - 1/(1+\tau)^2$. *Note the concavity of the reservation utility function. If we construct the compensation function with* α *and* β *in Proposition 1, we have*

$$E_x[C(r,t,x)] = -\frac{2}{(1+t)(1+\tau)} + 1 + \frac{1}{(1+t)^2}.$$

The first order condition still gives $t = \tau$, *but the second derivative of the expected compensation is*

$$\frac{\partial^2}{\partial t^2} E_x[C(r,t,x)] = -\frac{4}{(1+t)^3(1+\tau)} + \frac{6}{(1+t)^4}$$

which is positive at $t = \tau$, *violating the second order condition.* ■

The result of Proposition 1 holds only for non-concave reservation utility functions. When the reservation function is sufficiently concave, the compensation scheme derived from the first order condition becomes convex, barring the expert from revealing his own type to maximize compensation. To achieve the first best outcome

---

constant reservation utility. From (1.7), the participation constraint is binding for the lowest type, i.e., $\beta(\underline{\tau}) - \alpha(\underline{\tau})/(\tau_x + \underline{\tau}) = w$.

Then, the principal should solve

$$\min_{\alpha(\tau)} E[c(\tau)] = \int_{\underline{\tau}}^{\overline{\tau}} \left(w + \frac{\alpha(s)}{(\tau_x + s)^2} \frac{1 - F(\tau)}{f(\tau)} \right) f(\tau) d\tau$$

Point-wise minimization gives $\alpha(\tau) = 0$ for all τ and consequently $\beta(t) = w$ for all τ. That is, the optimal contract indicates that the principal offers flat wage.

The problem in this case is that the honest reporting and the truthful type revelation are implemented only weakly: experts are indifferent between sending truthful messages and lying. However, the principal can achieve the first best with arbitrary small cost by setting $\alpha(t)$ to be increasing in t very slowly but still keeping $\beta(t) = w$.

with a concave reservation utility function, the principal needs to make the compensation function more concave in equilibrium. This can be done by restricting the expected payoff to be linear in a geometric power of variance. We state this result in Proposition 2.

PROPOSITION 1.2: *Suppose $\underline{u}$ is a strictly increasing $\mathcal{C}^2$ function and the support of types is bounded, i.e. $\overline{\tau} < \infty$. Then, there exist $p \in \mathbb{N}$ such that the compensation scheme*

$$C(r, t, x) = -\alpha(t)(r - x)^{2p} + \beta(t)$$

achieves the first best outcome, where

$$\alpha(t) = \frac{2^p (p-1)!}{(2p)!} (\tau_x + t)^{p+1} \underline{u}'(t)$$

and

$$\beta(t) = \frac{(\tau_x + t)}{p} \underline{u}'(t) + \underline{u}(t).$$

PROOF. Let α and β be $\mathcal{C}^2$ functions on $\mathbb{R}_{++}$. Let $\alpha(t) \geq 0$ to satisfy (ICR). Then, the expected payoff for type τ is

$$E_x \left[C(r, t, x) \right] = -\alpha(t) \mu_{2p}(\tau) + \beta(t)$$

where $\mu_{2p}(\tau)$ is the $(2p)$'th central moment. Under the Gaussian specification, we have

$$\mu_{2p}(\tau) = E(r - x)^{2p} = \frac{(2p)!}{2^p p!} \left(\frac{1}{\tau_x + \tau} \right)^p$$

The first order condition for (ICT) is

$$\forall \tau \in [\underline{\tau}, \overline{\tau}], \;\; -\alpha'(\tau)\mu_{2p}(\tau) + \beta'(\tau) = 0 \tag{1.11}$$

Defining $\mathbf{C}(\tau, t)$ as in proposition 1, we have

$$\frac{\partial \mathbf{C}}{\partial t}(\tau, t) = -\alpha'(t) \cdot \frac{t - \tau}{(\tau_x + t)} \mu_{2p}(\tau),$$

which implies that $t = \tau$ is the global maximizer of $\mathbf{C}(\tau, t)$ if and only if $\alpha(t)$ is nondecreasing.

The expected payoff at the optimum, $c(\tau)$, is now $c(\tau) = \mathbf{C}(\tau, \tau) = -\alpha(\tau)\mu_{2p}(\tau) + \beta(\tau)$. From (1.11), we get

$$\begin{aligned}\beta(\tau) &= \beta(\underline{\tau}) + \int_{\underline{\tau}}^{\tau} \alpha'(s)\mu_{2p}(s)ds \\ &= \beta(\underline{\tau}) - \alpha(\underline{\tau})\mu_{2p}(\underline{\tau}) + \alpha(\tau)\mu_{2p}(\tau) - \int_{\underline{\tau}}^{\tau} \alpha(s)\mu'_{2p}(s)ds \end{aligned} \tag{1.12}$$

and

$$c(\tau) = -\alpha(\tau)\mu_{2p}(\tau) + \beta(\tau) = \beta(\underline{\tau}) - \alpha(\underline{\tau})\mu_{2p}(\underline{\tau}) - \int_{\underline{\tau}}^{\tau} \alpha(s)\mu'_{2p}(s)ds.$$

The principal's problem is now

$$\min_{\alpha(\tau)} E_\tau\left[c(\tau)\right] = \int_{\underline{\tau}}^{\overline{\tau}} \left(\beta(\underline{\tau}) - \alpha(\underline{\tau})\mu'_{2p}(\underline{\tau}) - \alpha(\tau)\mu'_{2p}(\tau)\frac{1-F(\tau)}{f(\tau)}\right) f(\tau)d\tau. \tag{1.13}$$

Note that the virtual cost in (1.13) still keeps the linearity in α, which implies the participation constraint should bind for all τ in the optimal contract, i.e.,

$$c(\tau) = \beta(\underline{\tau}) - \alpha(\underline{\tau})\mu_{2p}(\underline{\tau}) - \int_{\underline{\tau}}^{\tau} \alpha(s)\mu'_{2p}(s)ds = \underline{u}(\tau). \tag{1.14}$$

Differentiating (1.14) with respect to τ, we get

$$\alpha(t) = -\frac{\underline{u}'(t)}{\mu'_{2p}(t)}$$

and from (1.12) and the appropriate boundary condition,

$$\beta(t) = -\frac{\mu_{2p}(t)}{\mu'_{2p}(t)}\underline{u}'(t) + \underline{u}(t).$$

The remaining part is to check the second order condition or monotonicity of α. Since $\mu'_{2p}(t) = -pL(\tau_x + t)^{-p-1} < 0$ and $\mu''_{2p}(t) = -p(-p-1)L(\tau_x + t)^{-p-2}$ where $L = (2p)!/(2^p p!)$, we have

$$\alpha'(t) = -\frac{\underline{u}''(t)\mu'_{2p}(t) - \underline{u}'(t)\mu''_{2p}(t)}{\left(\mu'_{2p}(t)\right)^2} > 0$$

or

$$\frac{\underline{u}''(t)}{\underline{u}'(t)} > \frac{\mu_{2p}''(t)}{\mu_{2p}'(t)} = \frac{-p-1}{\tau_x + t}. \tag{1.15}$$

Since t is defined on a compact set and $\underline{u}'$ and $\underline{u}''$ are continuous, the left side of (1.15) is bounded. Therefore, for large p, the inequality holds for all t in the support of τ. ■

The logic of proposition 2 is as follows. To satisfy the second order condition, the sorting variable $\alpha(t)$ must be monotone increasing.[12] The $\alpha(t)$ derived from the first order condition is the product of $\underline{u}'(t)$ and a function of the announced posterior precision, which we call here $h(\tau_x{+}t)$. In the proposition, $h(\tau_x{+}t) = M(\tau_x{+}t)^{p+1}$ for a constant M. Though h turns out to be increasing and positive, α is not guaranteed to be monotone increasing for a concave $\underline{u}$. The principal, however, can take arbitrarily large p so that h increases fast enough to cover the effect of decreasing $\underline{u}'(t)$ so that the product is monotone increasing. Indeed, for a given reservation utility, we can find p^* such that any payoff function with $p > p^*$ achieves the first best. We present an example.

EXAMPLE 1.2: *Let $\tau_x = 1$ and $\underline{u}(\tau) = 1 - 1/(1+\tau)^2$, as in the second case in example 1. Then, from proposition 2, the expected payoff given honest reporting on the true state is, in the optimal contract with $p = 3$,*

$$E_x\left[C(r,t,x)\right] = -\alpha(t)\mu_{2p}(\tau) + \beta(t) = -\frac{2}{3}\frac{1+t}{(1+\tau)^3} + 1 - \frac{1}{3(1+t)^2}.$$

The first order condition gives

$$\frac{2}{3(1+\tau)^3} - \frac{2}{3(1+t)^2} = 0 \quad \Rightarrow \quad t = \tau$$

and the second order condition is satisfied since

$$\frac{\partial^2}{\partial t^2} E_x\left[C(r,t,x)\right] = -\frac{2}{(1+t)^4} < 0. \quad ■$$

[12]This is indeed equivalent to the supermodularity of the objective function in (t, τ).

Figure 1.1 shows how the power of the ex-post error affects the expected payoff. If the compensation is linear in mean squared error ($p = 1$), the truthful report $t = \tau = 1$ does not maximize the expected payoff. When the performance measure is more sensitive to the error, for example with $p = 3$ in this example, the truthful report becomes optimal for the expert.

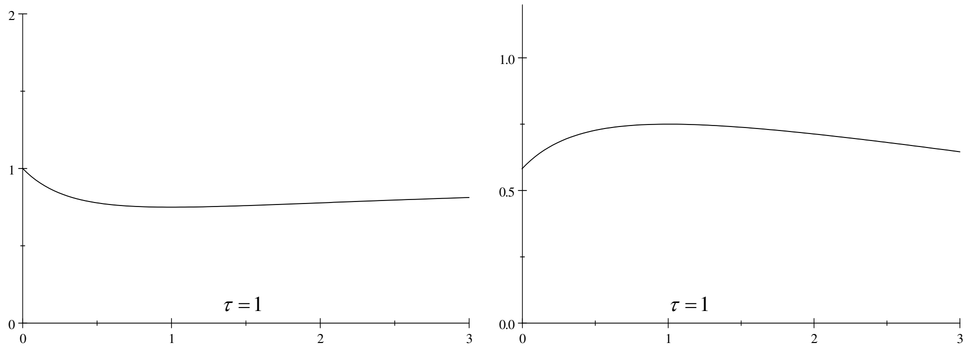

Figure 1.1: Expected Payoffs when $p = 1$ (right) and $p = 3$ (left)

The results in this section are interesting from two perspectives. First, sorting and honest reporting are implemented through a simple linear-form payoff function. This is because the linear payoff function under honest reporting satisfies the Spence-Mirrlees condition, or $-\alpha(t)/(\tau_x + \tau)$ is supermodular in α and τ. Moreover, it is supermodular in t and τ provided α is increasing in t. This simplifies the problem since the second order condition is equivalent to the monotonicity of α.

Another striking result is that the minimal information rent is paid in equilibrium. This is because, unlike the standard screening problem, the professional advising has a cheap talk feature in the sense that the sorting variable α does not affect the intrinsic cost or utility of the expert. This makes the virtual surplus (or virtual cost) linear in α. Therefore, the principal can fully control the payoff so the participation constraint is binding for all types of experts.

It is worthwhile noting that this mechanism is not a unique. One may design other mechanisms that achieve truthful type revelation and honest reporting. In addition, we should emphasize the compensation of each expert depends only on each expert's own report, not others'. This *independent* compensation scheme, in conjunction with the binding participate constraint, is beneficial to the principal because she can design the employment policy independent of the compensation scheme. We now turn our focus to the employment stage.

1.5 Optimal Employment

In the previous section, we showed that the principal can elicit each expert's precision and induce the honest reporting through a compensation scheme. Those features do not change when the principal wishes to hire more than one expert, since the optimal compensation scheme proposed is independent. Each expert would not care what types of experts he will co-work. Furthermore, since each expert will be paid at his reservation utility level, he would not concern about whether he will be hired or not. This implies that once pre-screening is done before the employment decision is made, the employment policy can be independent of the compensation scheme.

Specifically, consider the following game. At the beginning of the game, the true state is realized, but not revealed to anyone in the game. Then the pre-screening stage begins. The principal announces the compensation scheme, which is designed to screen the type of each applicant. Each expert, drawn from the population, applies for the job positions and send a message t on his own type. In the employment stage, the principal decides which applicants he will hire, based on the information he learns from the pre-screening stage. Once hired, each expert submits his report on the true state. Finally, the true state is revealed and payoffs are made according to the compensation scheme.

The screening through compensation simplifies the optimization problem. After the principal pre-screens experts, she knows the type of each applicant and how much she should pay if she hires some of them. Since experts hired are expected to submit honest reports on the true state, the objective function of the principal becomes a function of the precisions of employed experts less the sum of their reservation utilities. With the mean squared error specification of the revenue, for example, we have the optimization problem as follows:

$$\max_{S \subset I} \left(-\frac{a}{\tau_x + \sum_{j \in S} \tau_j} \right) - \sum_{j \in S} u(\tau_j) \tag{1.16}$$

where I is the set of all applicants. Now, the optimization problem becomes a combinatorial optimization, or a discrete portfolio problem, which is covered in Chapter 2 of this dissertation.

1.6 Conclusion and Discussion

This paper studies issues involving a principal wishing to hire possibly multiple experts for advice. To aggregate information from multiple sources and pay minimum amount to each type of expert, the principal needs to design a mechanism induces truthful type revelation and honest reporting on the true state. Under a Gaussian specification, it is shown that there exists a payoff function which achieves this first-best outcome. In the optimal contract proposed, the penalty for incorrect report is increasing in type (precision) revealed by experts, preventing less precise experts to hide behind more precise experts.

We derive the optimal compensation scheme in the class of linear functions in a specified performance measure. We show that if the reservation utility is convex, the first-best outcome is achieved with a payoff function linear in mean squared error. In the case with concave reservation utility, the performance measure should be more sensitive to the ex-post error, but still we can design the payoff function which is linear in the power of mean squared error.

In the paper, we assume each expert's gain from providing information depends only on the *current period* compensation paid by the principal, and his ability is elicited through it. However, from a dynamic perspective, the ability or precision of each agent may be evaluated by two parties: the decision maker (the principal) and the outside evaluator (the market). In this case, the gain from information provisions would come, at least partly, from future payoffs which depend on the reputation built on today. Recent empirical studies show that career concerns matter in expert advising. Ottaviani and Sørenson (2006a, 2006b) explore the theoretical approaches on this topic.

However, Ottaviani and Sørenson (2006a) assume the compensation is solely determined by reputation. In this sense, the approach of the paper is in the opposite direction to ours. It would be a complete theory only when we consider the compensation determined by both factors: future payoff from reputation and current payoff from compensation. As two parties are involved in evaluation, there would be a conflict of interests between the decision maker and the evaluator. Since the report tends to be shaded or biased in the presence of reputation concerns, the principal's objective is to reduce such effect, without paying too much. We leave these topics for future research.

Chapter 2

Getting Advice from Experts: Optimal Employment

2.1 Introduction

People seek advice from better informed people, or experts. Corporate CEOs often make decisions based on reports submitted by informed employees or sometimes from outside consultants. Governments operate and/or keep close contact with research institutes whose main objective is to advise clients of policies based on correct and accurate information. Stock market analysts help traders make best decisions by enlarging their information. In every situation mentioned above, the decision maker may need advice from more than one expert.

This chapter[1] explores the optimal employment when a monopolistic information demander, a principal, wishes to gather information from multiple experts. We assume each expert has heterogeneous information quality, or precision of the signal, and his reservation wage depends on the quality. When the information quality is known to the principal, his objective is to select the set of experts which provides the best information.

Even when the information qualities are private, the principal needs to elicit them to achieve the efficient information aggregation and minimal type dependent wage payment. In Chapter 1, we show that there exist compensation schemes with which the principal can screen the precision of each expert's signal and pays only the

[1]This chapter represents joint work with Jooyong Jun.

reservation utility. Through the proposed compensation scheme as a pre-screening device, the employment problem the principal faces becomes the same as one under full information on the precision.

We assume the principal's objective is to minimize the expected mean squared error of her own prediction on the true state less the wage payments. Under the Gaussian specification, the expected mean squared error is the principal's posterior variance, or the reciprocal of the posterior precision. The objective function of the principal's optimization problem is the set function, but with the assumptions and specifications of the model, it is transformed to the function of the sum of precisions.

With the nonlinear objective function and an arbitrary reservation utility function, it is hard to characterize the optimal employment set in general. We could, however, find interesting results under a specification on reservation utility. When the reservation utility is proportional to the marginal single information contribution, we show that the optimal employment set satisfies a cut-off property where experts with higher precision are hired. This strong result is derived because the marginal information contribution function becomes less concave than the reservation wage function as information is accumulated. We also discuss the case in which the cut-off rule does not hold, and relate the problem with the 'combinatorial optimization problem,' recently introduced to literature in economics.

As far as we know, there has been no paper which studies multiple experts and associated employment decision problem. However, the problem is analogous to the newly developed combinatorial optimization technique. In economics, Kelso and Crawford (1982) pioneered the technique, first introducing the 'gross substitutes condition' which guarantees a greedy algorithm to be optimal. The technique has since been used in cooperative game theory, matching models, and multi-item auctions. For recent literature, refer to Gul and Stacchetti (1999) which develops equivalent conditions with gross substitute condition and relates it to the Walrasian equilibrium, Lehmann et. al. (2006) and Milgrom and Strulovici (2006) for auctions, and Chade and Smith (2005) which develops a greedy algorithm to solve a simultaneous search problem with single stochastic prize without gross substitutes condition.

This chapter is organized as follows: We describe the model and present the optimization problem in the next section. In section 3, we present the main result: under a specification of the reservation utility function, we derive the property of the optimal employment set. Section 4 generalizes the result and propose the conditions under which the main result holds. Section 5 relates the result with the recent literature on combinatorial optimization. Section 5 concludes and addresses issues for further research.

2.2 Model

An uninformed principal tries to make the best prediction of the true state, for example the profitability of a project. To get better information beyond common prior, the principal wishes to hire privately informed agents, who are called 'experts' hereafter. Experts are heterogeneous in the precision of their private signals on the true state, which is labelled their *type*.

The true state x is drawn from a normal distribution with mean μ_x and variance $1/\tau_x$, which is common knowledge. We assume the true state is verifiable and thus contractible. While the principal has no private information, expert $i \in I$ receives a conditionally independent private signal s_i, where I is the set of experts who applied for the job positions. The distribution of s_i is assumed to follow a normal distribution with mean x and precision τ_i, or

$$s_i = x + \epsilon_i, \quad \epsilon_i \sim N\left(0, 1/\tau_i\right).$$

The precision, or type, of each expert τ_i is drawn from the population with distribution function F on the support of $[\underline{\tau}, \overline{\tau}] \subset \mathbb{R}^+$. We assume F is continuously differentiable so that the probability continuous density function f exists. The type of each expert is assumed to be either common knowledge, or known to the principal through costless pre-screening before the principal makes the employment decision.

Once hired, each expert should submit a *report* on the true state, denoted by $r_i \in \mathbb{R}$ for expert i. We assume the compensation scheme is designed for each expert to report his posterior mean. In other words, the honest report is a priori assumed. Since the principal knows the precision of each expert, the honest report on the true state is equivalent to the honest report on the signal received. Formally, we have the one-to-one relationship between the private signal and the honest report:

$$s_i = \frac{(\tau_x + \tau_i)r_i - \tau_x \mu_x}{\tau_i}$$

Experts are assumed to be risk neutral utility maximizers with the identical vNM utility function $u(c) = c$. Each expert has a reservation utility which is type-dependent. With risk neutrality, the reservation utility can be seen as the reservation wage. The reservation wage, or the outside option might be from the experts' labor market, or from the single production through the private information. We focus on the second interpretation, noting that the principal is a monopolistic information demander. We will discuss the first interpretation in later section. Type-dependent reservation utility function $\underline{u}(\tau)$ is assumed to be strictly increasing and continuously differentiable.

The principal's objective is to maximize *revenue* less payoffs to employed experts. The revenue function R depends on the principal's prediction on the true state, denoted by $\widehat{x}$, and the true state. We specify it as $R(x, \widehat{x}) \equiv K - a(\widehat{x} - x)^2$ for constants $K > 0$ and $a > 0$. I.e., the principal tries to minimize the mean squared error[2].

[2]The qualitative results of this chapter hold provided the principal minimizes any power function of ex-post error, $| \hat{x} - x |$. This is because the expectation of the power of error is a constant times the power of the posterior variance. One can easily transform the optimization problem into one with the mean squared error.

With known precisions and honest reports, the principal's best prediction is the posterior mean of the true state given all information available for her. If $J \subset I$ is the employment set, we have

$$\widehat{x} = \frac{\sum_{j \in J} r_j(t_j + \tau_x) - (|J| - 1)\mu_x \tau_x}{\sum_{j \in J} t_j + \tau_x},$$

This prediction is unbiased, so the expected mean squared error becomes the posterior variance, $\tau_x + \sum_{j \in J} \tau_j$. Principal's objective is to select the best subset $S \subset I$ which maximizes the revenue less the wage payments. The optimization problem is formally described as follows.

$$\max_{S \subset I} E_x \left[K - a(\widehat{x} - x)^2 - \sum_{j \in J} \underline{u}(t_j) \right] = \left(K - \frac{a}{\tau_x + \sum_{j \in S} \tau_j} \right) - \sum_{j \in S} \underline{u}(\tau_j) \qquad (2.1)$$

In principle, the problem can be solved through computing values of the objective function over the power set of alternatives (in our setting over the power set of applicants). It belongs to the class of *combinatorial optimization problems*, which aims at finding the best subset from a finite set of alternatives. Indeed, the optimization problem is a special case of a firm's employment decision problem in Kelso and Crawford (1982) even though the focus is quite different from ours.[3]

In our model, we can consider the expected information gain of the principal as a joint production by employed experts. As will be shown in the next section, the joint production set function is submodular with respect to the set inclusion operation. This implies that the production function exhibits the decreasing marginal returns: the marginal contribution by a single agent to a subset of alternatives decreases as the subset becomes larger. Though the submodularity is a nice property for the objective function in maximization problems, it does not guarantee that we have a simple algorithm to get the solution. Indeed, it is well known that the maximization

[3]The utility maximization problem given a price vector in Gull and Stacchetti (1999) is also isomorphic to ours, where the utility and the price are analogue to the information gain and the reservation wage, respectively.

of a general submodular set function is computationally intractable[4]. This makes it difficult not only to find the efficient algorithm but also to characterize the property of the optimal set.

Nonetheless, the objective function of our model has a nice feature. Any set of experts can be characterized by a single real number, the sum of precisions of experts in the set. This allows us to transform the objective set function to a function on the two dimensional Euclidian space. This transformation allows us to solve the problem through a greedy algorithm, as will be discussed later.

Yet the optimal employment set is quite arbitrary since it depends heavily on the form of reservation utility function. We make a critical but reasonable assumption on the reservation utility in the next section: the reservation utility is proportional to the marginal single information contribution. Under this specification, the optimal employment set is shown to follow a cut-off property. We then discuss on general cases, providing an example of complicated optimal employment set. The comparison of our model and other combinatorial optimization problem is presented in the final subsection.

2.3 Properties of Optimal Employment Set

Before we begin the analysis, we define some functions for notational convenience.

Definition 2.3.1: (Objective function)

1) A set function g, called *information gain function* hereafter, is a mapping from the power set of I to $\mathbb{R}_+$, which is defined as

$$g(S) \equiv K - \frac{a}{\tau_x + \sum_{j \in S} \tau_j}, \quad S \subset I.$$

[4]In the combinatorial optimization problem theory, the problem is known to be so called NP-hard: there is no algorithm for the problem to be solved in a polynomial time. See, for example, Lovàsz (1982).

If S is a singleton, $g(S)$ is called a *single information gain function.* We abuse notation by denoting $g(i) = g(\{i\})$ for $i \in I$.

2) We call f the *information contribution (IC) function*, denoted by $f(T_S)$ where T is the sum of the precisions in the set $S \subset I$[5];

$$f(T) \equiv K - \frac{a}{T}.$$

Note that $T \in [\tau_x, \tau_x + \sum_{i \in I} \tau_i]$.

3) *The marginal information contribution (MIC) function* is derived from information contribution function. For prior T and precision of a new signal τ, we let

$$\triangle\,(T, \tau) \equiv f(T + \tau) - f(T) = \frac{a\tau}{T(T + \tau)}.$$

Similarly we define the information loss when we remove a signal with precision τ from a set whose precision is T.

$$\nabla(T, \tau) \equiv f(T) - f(T - \tau) = \frac{a\tau}{T(T - \tau)}.$$

We have the following properties of the information contribution function.

LEMMA 2.1: *(Properties of IC function)*

(a) *(Monotonicity of information gain function) For any* $A \subset B \subset I$, $g(A) \leq g(B)$.

(b) *(Monotonicity of information contribution function)* $f(T)$ *is increasing in* T.

(c) *(Submodularity of information gain function) For any* A *and* B *in* 2^I, *we have*

$$g(A) + g(B) \geq g(A \cup B) + g(A \cap B).$$

(d) *(Concavity of IC function) For any* $\tau_1 > \tau_2$ *and for any* $T_1 > T_2$, *the inequality* $f(T_1 + \tau_1) - f(T_1 + \tau_2) < f(T_2 + \tau_1) - f(T_2 + \tau_2)$ *is satisfied.*

[5] We sometimes suppress S.

Proof. It is easy to show that (a) and (b) are equivalent and (c) and (d) are equivalent. (b) is obvious by definition. To see (c) holds, it suffices to show that

$$g(A) - g(A \cap B) \geq g(A \cup B) - g(B)$$

$$\Leftrightarrow$$

$$\frac{1}{\tau_x + \sum_{j \in A \cap B} \tau_j} - \frac{1}{\tau_x + \sum_{j \in A} \tau_j} \geq \frac{1}{\tau_x + \sum_{j \in B} \tau_j} - \frac{1}{\tau_x + \sum_{j \in A \cup B} \tau_j}.$$

This inequality holds since

$$\frac{\sum_{j \in A \setminus B} \tau_j}{(\tau_x + \sum_{j \in A \cap B} \tau_j)(\tau_x + \sum_{j \in A} \tau_j)} \geq \frac{\sum_{j \in A \setminus B} \tau_j}{(\tau_x + \sum_{j \in A \cup B} \tau_j)(\tau_x + \sum_{j \in B} \tau_j)}.$$

■

We now make a critical assumption on the reservation utility. We assume that it is proportional to the marginal single information contribution[6], that is, for $k > 1$,

$$\underline{u}(\tau) \equiv \frac{1}{k} \triangle (\tau_x, \tau) = \frac{a\tau}{k\tau_x(\tau_x + \tau)}.$$

We also normalize $a = \tau_x = 1$. The assumption is quite strict but reasonable. It says that if the expert utilizes his private information outside of the relationship with the principal, his gain is proportional to the marginal single information gain. Behind the assumption we think all potential principals in the market share the same information on each expert so that the gain inside of the market should be the same across principals. Since all employers do not know who is what type, the reservation utility cannot be type dependent. The only situation where an expert with higher ability gains more lies in the case when he uses the private information for his own gain.

To prove the key characterization of the optimal employment set, we need the following lemma:

[6]We strongly conjecture that the main result still hold if the reservation utility is a concave function of the marginal single information contribution. This is because, as will be clarified later, the main result depends on the fact that the marginal contribution function near the global optimum crosses the reservation utility only once and from below. The fact is still satisfied when the reservation utility is concave in the marginal single contribution.

Lemma 2.2: *The net information gain from adding* τ *from a set with collective precision* T, $\triangle\ (T,\tau) - \underline{u}(\tau)$ *crosses zero on* $\tau > 0$ *at most once and from below. Likewise, the net information gain from dropping* τ, $\underline{u}(\tau) - \nabla(T,\tau)$ *crosses zero at most once and from above.*

Proof. $\triangle\ (T,\tau) - \underline{u}(\tau) > 0$ if and only if $T(T+\tau) - k(1+\tau) < 0$. Since it is linear in τ, for some $\tau > 0$ to satisfy the equality we must have either $k - T^2 > 0$ and $T - k > 0$ or $k - T^2 < 0$ and $T < k$. However, the latter inequality cannot hold because $k > 1$. Thus, we need to check only the case of $\sqrt{k} < T < k$. Then, $T(T+\tau) - k(1+\tau)$ is a decreasing function of τ and crosses zero only once rom above. This implies that $\triangle\ (T,\tau) - \underline{u}(\tau)$ crosses zero at most once from below.

For the dropping case, $\underline{u}(\tau) - \nabla(T,\tau) > 0$ if and only if $(T+k)\tau - T^2 + k < 0$. This crosses zero at most once regardless of the value of T. Since it crosses from below, $\underline{u}(\tau) - \nabla(T,\tau)$ crosses from above. ■

The intuition of Lemma 2 is as follows. For the marginal information contribution function to cross the reservation utility function, T must be in an appropriate range. Since the MIC becomes less concave as T increases, it is flat relative to the reservation wage function in the range of T.

We need an additional lemma to prove the main proposition.

Lemma 2.3: *Suppose that* $T^2 > k$. *Let* τ_2 *satisfy* $\triangle\ (T,\tau) - \underline{u}(\tau) = 0$ *and* τ_1 *satisfy* $\underline{u}(\tau) - \nabla(T,\tau) = 0$. *Then,* $\tau_1 < \tau_2$.

Proof. The existences of τ_1 and τ_2 are immediate from the proof of lemma 2. We have $(T-k)\tau_2 + T^2 - k = 0$ and $(T+k)\tau_1 - T^2 + k = 0$. But then, $(T+k)\tau_2 - T^2 + k > -(T-k)\tau_2 = T^2 - k$. Thus, $\underline{u}(\tau_2) - \nabla(T,\tau_1) < 0$, which implies that $\tau_1 < \tau_2$ ■

Proposition 2.1: *Under the current reservation utility specification, the optimal employment set follows the cut-off property. That is, there exists* τ^* *such that experts* i *is in the optimal set if and only if* $\tau_i \geq \tau^*$.

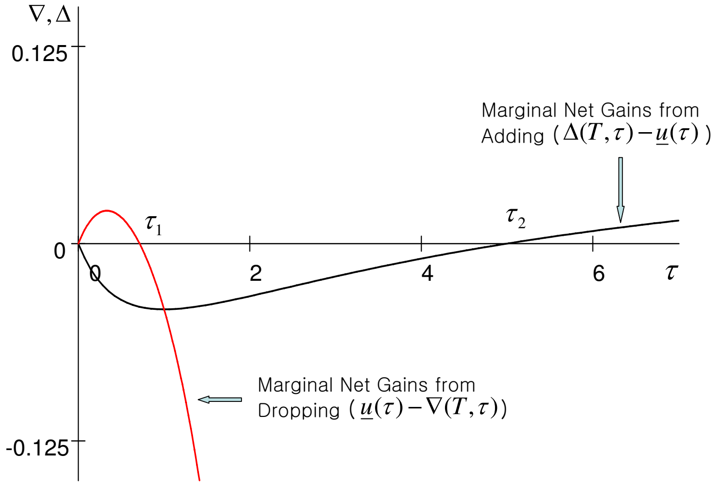

Figure 2.1: Marginal Gain from Adding/Dropping an Expert with τ ($k = 4$, $T = 3$)

Proof. Let $S \subset I$ be the optimal set and let T_S be the associated collective precision. Note first that if $T_S^2 < k$, S cannot be the optimum unless $S = I$ since adding any expert in $I \setminus S$ yields positive net gain. We only consider the case $T_S^2 > k$.

Define τ_1 and τ_2 as in the proof of lemma 3. Suppose $s \in S$ is less than τ_1. Then, dropping it improves net gain, contradicting the optimality. Similarly, any $u \in I \setminus S$ cannot be bigger than τ_2. The only thing we need to check is the case in which there are i and j such that both are between τ_1 and τ_2, $\tau_i < \tau_j$, and $\tau_i \in S$ but $\tau_j \in I \setminus S$. Consider $S \setminus \{\tau_i, \tau_j\}$. The optimality implies that $\triangle\ (T_S - \tau_i, \tau_i) - \underline{u}(\tau_i) > \triangle\ (T_S - \tau_i, \tau_j) - \underline{u}(\tau_j)$. But then $\triangle\ (T - \tau_i, \tau) - \underline{u}(\tau)$ crosses zero from above, which contradicts Lemma 2. (Refer to Figure 1) This completes the proof. ■

It should be noticed that the cut-off property is a property of the optimal employment set, not a decision rule. A decision rule is a method to find the optimal set given the knowledge of the applicants' types from the pre-screening stage. Even though the cut-off type τ^* depends on the types of applicants, it does not mean that we can specify a decision rule as a function $\tau^*(\tau_1, \cdots, \tau_{|I|})$.

However, the cut-off property suggests a simple algorithm work to find the optimal set. We propose two algorithms. The first one is the *Marginal Improvement Algorithm (MIA)*, as in Chade and Smith (2005). The principal begins with the null set and search the best expert and add him into the employment set. She searches the best expert given the current employment set (and thus the current T) among experts not in the employment set. She repeats this procedure until the marginal net information gain is negative. The second algorithm is as follows. The principal first sorts the applicants by type, then begins checking starting with the highest type of applicant. She adds the applicant to the employment set as long as the marginal net gain is positive. After all, she hires all applicants before the applicant whose marginal net gain is negative. Both algorithms solve the problem in $\mathcal{O}(|I|^2)$, steps but they are different in orders of experts added to the employment set.

The following example shows the optimal employment set when $|I| = 2$.

EXAMPLE 2.1: *Consider the two applicants case:* $|I| = 2$. *Let* $a = 1$, $t_x = 1$, *and* $t_i \in (0, 2]$. *Also, assume that the reservation utility for the expert with type* τ *is half of his marginal single production. That is,*

$$\underline{u}(\tau) = \frac{1}{2}\left(g(\{\tau\}) - g(\emptyset)\right) = \frac{\tau}{2\tau_x(\tau_x + \tau)}.$$

The graph shows the outcome of the optimal employment decision. We have

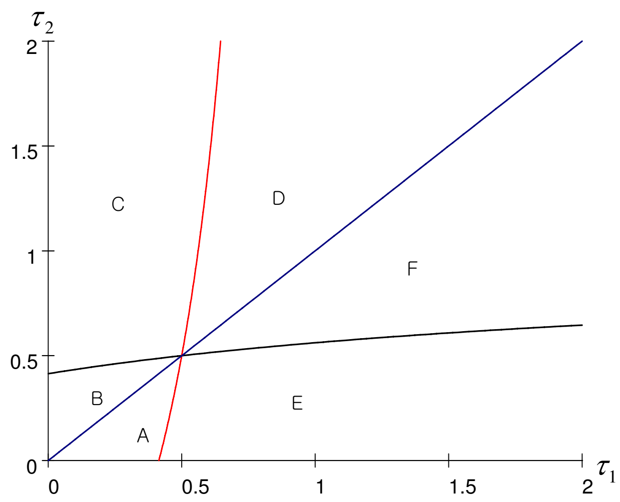

Figure 2.2: Optimal Employment Set for Two Applicants

Area	*Net Gain for the Principal*
A	$\{1,2\} \succ \{1\} \succ \{2\}$
B	$\{1,2\} \succ \{2\} \succ \{1\}$
C	$\{2\} \succ \{1,2\} \succ \{1\}$
D	$\{2\} \succ \{1\} \succ \{1,2\}$
E	$\{1\} \succ \{1,2\} \succ \{2\}$
F	$\{1\} \succ \{2\} \succ \{1,2\}$

We can easily check the monotonicity of the optimal employment set: there is no case in which the higher type is not employed while the lower type is. This is equivalent to the cut-off property of the optimal set. It shows also that both applicants are hired together only when they are all of low type. ■

2.4 Discussion on the Specification

2.4.1 Other Reservation Utility Specification

In the previous section, we show a nice property for the optimal employment set. The result crucially depends on the quasi-convexity of the marginal net information gain, which is due to the fact that, roughly speaking, the marginal information contribution is less convex than the reservation wage. Moreover, the marginal information contribution becomes less concave as T gets large. This implies that even though the marginal information contribution is more concave than reservation utility at the initial state (when $T = \tau$), it might become less concave when T approaches the optimal cut-off point. The specification of reservation utility in the previous subsection shows exactly this case. Initially, the curvature of the reservation utility is the same as that of the marginal information contribution. For $T > \tau_x$, however, the curvature of the former is always bigger than the latter.

Though the assumption on reservation utility in the previous subsection is reasonable, there are other possibilities where the property we discussed above does not hold. To motivate, we provide an example in which the simple greedy algorithm described above fails to lead to the optimum.

EXAMPLE 2.2: *Figure 3 shows a linear reservation utility function* $\underline{u}(\tau) = \frac{1}{16}\tau$ *and marginal information contribution functions when* $a = 1$, T *being varied. Let* $\tau_x = 1$, *and the support of* τ *is* $(0, 3]$. *Note first that when the principal is to hire only one expert, the highest type is always preferred since* $\triangle\,(1, \tau) \equiv f(1, \tau) - \underline{u}(\tau)$ *is increasing when* $\tau \leq 3$.

Suppose three applicants have applied for the job and from the pre-screening the principal knows their types are $\tau_1 = 1$, $\tau_2 = 2$, *and* $\tau_3 = 2.8$. *The global optimum is to hire expert 1 and expert 2. However, if the principal solves locally, he will first hire expert 3, and then will not hire any more expert, as demonstrated in Figure 3.*

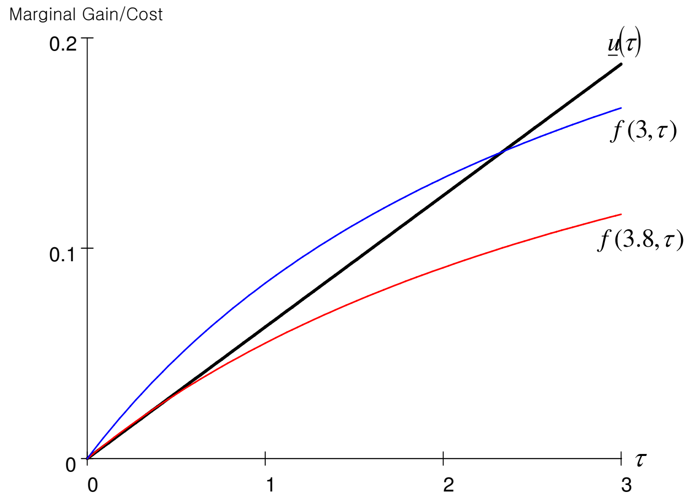

Figure 2.3: Marginal Information Gain and Reservation Utility

(Failure of MIA)

Consider now three applicants of type $\tau_1 = 1$*,* $\tau_2 = 1.5$*, and* $\tau_3 = 2$*. The global optimum is to hire expert 1 and expert 3. (Failure of monotonicity)* ■

The above example shows that the *marginal improvement algorithm* does not lead to the global optimal solution[7], and the optimal employment set may be non-monotonic. The key point in this example is the quasi-concavity of the marginal net gain. As the reservation utility function is close to linear, the marginal net gain is always more concave than the reservation utility function. If the reservation utility is linear or convex, the marginal net gain is always concave due to the submodularity of f. This implies that as the employment set is enlarged, or equivalently the information is cumulated, the lower type has a better chance of being hired than the higher type, though initially the higher type contributes more. This breaks down the monotonicity and the MIA may not lead to the global optimum.

[7]This does not mean that no greedy algorithm solves the problem. Greedy algorithm is a broader concept than MIA. It is, roughly speaking, a way to find the global optimal solution to a discrete optimization problem through finding local optima in each step. If such algorithm leads to the global optima, it reduces calculation time dramatically (from an exponential time to a polynomial time.)

2.4.2 General Conditions for the Cut-off Property

As we discussed in the previous section, the cut off property crucially depends on the fact that, near the global optimum, marginal information contribution crosses the reservation utility only once and from below. One may ask the general condition for the objective function under which the property holds. This subsection shows the single crossing property holds when the objective function satisfies the *decreasing curvature* property. This type of function is referred to as DARA for utility functions. We define this property in terms of discrete choices.

DEFINITION 2.4.1: A set function f on 2^S, where $S = \{s_1, \ldots s_n\}$, is said to satisfy decreasing curvature (DC, hereafter) if for a given , if for all $i = 1, 2, \cdots, n-2$,

$$-\frac{f(s_i) - 2f(s_{i+1}) + f(s_{i+2})}{f(s_i) - f(s_{i+1})}$$

is decreasing in i.

We first assume that the objective function f is a set function defined on 2^E where E is a finite set, and there exists a complete order among elements in E with respect to a binary relationship $\succeq$. Suppose, as in the main model, that the reservation utility $\underline{u}$ is proportional to the individual's single production, or $\underline{u}(a) = \alpha f(a)$ where $0 < \alpha < 1$. Assume $f(\emptyset) = 0$ so that the marginal single production is the single production. Then, for a set of applicants A, the maximum profit is $f(A) - \alpha \sum_{e \in A} f(e)$. The positive single crossing condition is satisfied if there exists $a^* \in S$ such that

$$f(a^* \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(a^*) \right) \geq 0$$

then, for any $b \succcurlyeq a^*$ and $b \notin A$,

$$f(b \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(b) \right) \geq 0$$

and for any $b \preccurlyeq a^*$,

$$f(b \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(b) \right) \leq 0$$

are satisfied.

Because f is an increasing, submodular set function, if the positive single crossing condition is satisfied, then for a subset $A' \supseteq A$, if there exists b^* such that for any $b \succcurlyeq b^*$,

$$f(b \cup A') - \alpha f(b) \geq \alpha \sum_{e \in A'} f(e)$$

is satisfied, we have $b^* \succcurlyeq a^*$.

The following proposition states that the positive single crossing condition is satisfied if the submodular production function shows a decreasing curvature and the reservation utility is given as a fraction of single production.

PROPOSITION 2.2: *For an increasing submodular set function $f(\cdot)$ and a subset $A \in 2^S$, if the reservation utility is given as a fraction of single production, or $\underline{u}(s) = \alpha f(s)$ where $s \in S$, $f(\cdot)$ satisfies the positive single crossing condition if $f(\cdot)$ satisfies decreasing curvature.*

PROOF. First, we explain the decreasing absolute risk averse property, a decreasing curvature, of a set function with a complete subset order for its domain. It also leads that if a non-empty subset A does not include s_i, s_{i+1}, nor s_{i+2}, then the inequality

$$-\frac{f(s_i \cup A) - 2f(s_{i+1} \cup A) + f(s_{i+2} \cup A)}{f(s_i \cup A) - f(s_{i+1} \cup A)} < -\frac{f(s_i) - 2f(s_{i+1}) + f(s_{i+2})}{f(s_i) - f(s_{i+1})}$$

is satisfied.

If there exists such n^*, the inequality

$$f(s_n \cup A) - f(s_{n^*} \cup A) \geq \alpha f(s_n) - \alpha f(s_{n^*}) \tag{2.2}$$

needs to be satisfied. From DC property, the following inequality

$$0 > f(s_n \cup A) - 2f(s_{n+1} \cup A) + f(s_{n+2} \cup A) > \alpha(f(s_n) - 2f(s_{n+1}) + f(s_{n+2})) \tag{2.3}$$

is satisfied, which implies that the decrease of production difference is slower than the decrease of reservation utility difference and, therefore, for any $s_n \succcurlyeq s_{n^*}$ (or

$n \leq n^*$) and $s_n \notin A$, (2.2) is always satisfied, which implies that $f(s_n \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(s_n) \right) \geq 0$.

If (2.2) is violated, from decreasing curvature, the decrease of reservation utility difference is slower than the decrease of production difference, which leads to the violation of (2.3). Then, there may exists s^* such that any s_n such that $s_n \succcurlyeq s^* \succcurlyeq s_{n^*}$ violates the inequality $f(s_n \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(s_n) \right) \geq 0$, which contradicts the definition of n^*.

Finally, we want to show that $f(s_n \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(s_n) \right) < 0$ for all $s_n \prec s_{n^*}$. Suppose there exists such $s' \notin A$ that $s' \prec s_{n^*}$ but $f(s' \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(s') \right) \geq 0$ and there exists s'' such that $s' \prec s'' \prec s_{n^*}$ and it satisfies $f(s'' \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(s'') \right) < 0$. Then, we have the following inequality $f(s'' \cup A) - \alpha f(s'') < f(s' \cup A) - \alpha f(s')$ which clearly violates (2.2) and from decreasing curvature, no $s \succ s'$ satisfies the inequality $f(s \cup A) - \alpha \left(\sum_{e \in A} f(e) + f(s) \right) \geq 0$. This result, however, clearly violate the assumption $s_{n^*} \succ s'$. This completes the proof. ∎

2.5 Efficient Algorithm

2.5.1 Gross Substitute and Efficient Algorithm

In an analysis of labor market equilibrium, Kelso and Crawford (1982) introduce the *Gross Substitutes (GS)* condition to show the existence and the stability of the equilibrium. Gul and Stacchetti (1999), in a slightly different setting, also consider the objective set function satisfying GS condition in addition to a new condition of a set function, the *Single Improvement (SI)* property. Though focuses of their papers are on the existence of market equilibrium, not on the optimal decision itself as ours, their results provide some implications for our model. The following definitions are ones slightly modified from Gul and Stacchetti (1999).

DEFINITION 2.5.1: Denote $S(w)$ be the optimal set of the optimization problem.

A function $f : 2^I \to \mathbb{R}$

(i) satisfies the GS condition if for any two reservation wage vectors w^1 and w^2 in $[\underline{w}, \overline{w}]^I$ such that $w^1 \leq w^2$, and any $A \in S(w^1)$, there exists $B \in S(w^2)$ such that $\{i \in A | w_i^1 = w_i^2\} \subset B$.

(ii) has the SI property if for any reservation wage vector w in $[\underline{w}, \overline{w}]^I$ and set of employees $A \notin S(w)$, there exists a bundle B such that $g(A) - \sum_{j \in A} w_j < g(B) - \sum_{j \in B} w_j$ and $|A \setminus B| \leq 1$, and $|B \setminus A| \leq 1$.

Gul and Stacchetti (1999) show that if f is monotone then the GS condition is equivalent to the SI condition. Chade and Smith (2006) states in the appendix that the *SI* condition, or equivalently $M^\natural$-concavity from Murota and Shioura (2003), guarantees that a greedy algorithm leads to the global optima. The greedy algorithm using SI condition, which is called the *steepest ascent algorithm (SAA)*, can be roughly as follows. Start with an arbitrary subset of the whole set. If the set is locally optimal, then stop. Otherwise, add or subtract a single element to/from the subset or replace one element in the set with an element from outside, which improves the objective in the best way. Repeat this procedure until we reach the local optimum.

MIA is a *special* case of the SAA in that the procedure does not contain a substraction nor a replacement in each local step. It solves the problem more efficiently than the SAA, but it is not guaranteed for the algorithm to lead the global optima even in the case when the objective function satisfies the SI condition. However, Proposition 3 implies that the principal in our model can solve the problem through the MIA, thus through the SAA.

In fact, the SI property holds for the objective function in our model since given a set of applicants, one can find the best among the applicants outside of the set and/or the worst among those inside the set, which is illustrated in Figure 1. This can be generalized even further. Once the objective function is a function of the sum of types of a subset, that is, if the set function can be transformed to a function whose domain is a Euclidean space, the SI property always holds for the objective function.

The SI property of the objective function in our model provides a weak comparative statics result. Due to the equivalence of the GS and SI conditions, the information gain function satisfies the GS condition. This implies that an increase in reservation wages for part of the applicant pool does not affect the employment decision on applicants whose reservation wages do not change.

2.5.2 Efficient Algorithm for a Submodular Set function

In the main model of this chapter, we consider a specific set function where it is a reciprocal of the sum of elements. This subsection considers the optimization of a general submodular set function. Though myopic local search does not lead to the global optima, we propose an algorithm from some properties of the submodular set function which reduces the calculation time.

Let $S_n := \{s_1, \ldots, s_n\}$, and $S_{n+1} := S_n \cup \{s_{n+1}\}$ for a given set of N applicants I. For sets A, B, and C, define $[A, B, C] \equiv (A \backslash B) \cup C$.

The following lemma states that once an element $x \in S$ is not included in the optimal subset of one set, x is not included in the optimal subset of any bigger set including the original set. It is a critical property which is used for the employment algorithm later.

LEMMA 2.4: *If $\tau \in S_n$ and $\tau \notin A_n$, then for any S_m such that $S_n \subset S_m \subseteq S$ (or for any $m > n$), $\tau \notin A_m$.*

PROOF. It is obvious that you cannot be better off by removing any element from an optimal subset or including any other element into it. In other words, for any $\tau \in A_n$, where A_n is the optimal subset for a given $S_n \subseteq S$, the inequality $f(A_n \backslash \{\tau\}) \leq f(A_n) - \underline{u}(\tau)$ must be satisfied, and for any $a \notin A_n$, $f(A_n \backslash \{a\}) \geq f(A_n) - \underline{u}(a)$. We also know that $f(A_n) \leq f(A_m)$.

Suppose there exists an element $\tau \notin A_n$ but $\tau \in A_m$. Then the inequalities

$f(A_n \cup \{\tau\}) - \underline{u}(\tau) < f(A_n)$ and $f(A_m \backslash \{\tau\}) < f(A_m) - \underline{u}(\tau)$ must be satisfied. From these inequalities, we know that $f(A_n \cup \{\tau\}) + f(A_m \backslash \{\tau\}) < f(A_n) + f(A_m)$ must be satisfied, which contradicts the submodularity assumption. Therefore, if $\tau \notin A_n$, then $\tau \notin A_m$. ■

The following lemma states that if including an additional element leads to the better payoff but not to an optimum, exchaging an element will do so.

LEMMA 2.5: *If* $f(A_n \cup \{s_{n+1}\}) - \underline{u}(s_{n+1}) > f(A_n)$ *and* $A_{n+1} \neq A_n \cup \{s_{n+1}\}$, *then there exists a unique* $a \in A_n$ *which satisfies* $A_{n+1} = [A_n, \{a\}, \{s_{n+1}\}]$.

PROOF. Note that $S_{n+1} \backslash S_n = \{s_{n+1}\}$. Lemma 1 implies that any element not in A_n cannot be included in A_{n+1}. Therefore, including any element $x \notin A_n \cup \{s_{n+1}\}$ does not leads to A_{n+1}. Therfore, there exists at least one $a \in A_n$ such that $f(A_n \cup \{s_{n+1}\}) - \underline{u}(a) \leq f([A_n, \{a\}, \{s_{n+1}\}])$ and the additional gain from hiring s_{n+1} is strictly positive.

Now we check uniqueness. Because $a \not\succcurlyeq \tau_{n+1}$, the following inequality

$$f([A_n, \{a\}, \{\tau_{n+1}\}]) \leq f(A_n)$$

must be satisfied, which also implies that $[A_n, \{a\}, \{\tau_{n+1}\}] \preccurlyeq A_n$. We already know that for any $\tau \in A_n$, the following inequality $f(A_n) - \underline{u}(\tau) \geq f(A_n \backslash \{\tau\})$ is satisfied. Therefore, with Assumption 1, the following inequality

$$f([A_n, \{a\}, \{\tau_{n+1}\}]) - f([A_n, \{a\}, \{\tau_{n+1}\}] \backslash \{\tau\}) \geq f(A_n) - f(A_n \backslash \{\tau\})$$

is satisfied for any $\tau \in [A_n, \{a\}, \{\tau_{n+1}\}]$.

Suppose now that there exists another element $a' \in A_n$ and $a' \neq a$ which satisfies $f([A_n, \{a\}, \{\tau_{n+1}\}]) - \underline{u}(a') \leq f([A_n, \{a\}, \{\tau_{n+1}\}] \backslash \{a'\})$. Then, from $f(A_n) - \underline{u}(a') \geq f(A_n \backslash \{a'\})$, $f([A_n, \{a\}, \{\tau_{n+1}\}]) + f(A_n \backslash \{a'\}) \leq f(A_n) + f([A_n, \{a\}, \{\tau_{n+1}\}] \backslash \{a'\})$ must be satisfied, which violates the submodularity assumption. Therefore, for any

$\tau \in [A_n, \{a\}, \{\tau_{n+1}\}]$, the following inequality

$$f([A_n, \{a\}, \{\tau_{n+1}\}]) - \underline{u}(\tau) > f([A_n, \{a\}, \{\tau_{n+1}\}] \backslash \{\tau\})$$

is always satisfied, which implies that there is no element to be removed other than a. Therefore, only one element in A_n needs to be removed to reach the new optimal subset A_{n+1} if $A_n \cup \{s_{n+1}\} \neq A_{n+1}$. ■

Here is the optimal algorithm of simultaneous hiring:

PROPOSITION 2.3: *The following algorithm finds the global optimum.*

(a) Sort the N applicants by the decreasing order of capability.

(b) If $f(A_n \cup \{\tau_{n+1}\}) - \underline{u}(\tau_{n+1}) < f(A_n)$, then set $A_{n+1} = A_n$ and move on to the next applicant.

(c) Otherwise, if $f(A_n \cup \{\tau_{n+1}\}) - \underline{u}(\tau_{n+1}) > f(A_n)$, then find $\underline{a} \in A_n$ which minimizes $f(A_n \cup \{\tau_{n+1}\}) - f([A_n, \{\underline{a}\}, \{\tau_{n+1}\}]) - \underline{u}(\underline{a})$. If this value is positive, then set $A_{n+1} = A_n \cup \{\tau_{n+1}\}$. If this value is negative, then set $A_{n+1} = [A_n, \{\underline{a}\}, \{\tau_{n+1}\}]$.

(d) Repeat 2-3 until reaching the final applicant.

2.6 Conclusion and Discussion

This paper studies issues involving a principal wishing to hire possibly multiple experts for advice. Based on the result in the pre-screening stage, we examine the optimal employment decision. We show that under a realistic specification on the reservation utility, a cut-off rule employment policy is optimal. We also provide an example in which the monotonicity and the marginal improvement algorithm break down. It turns out that the property of the optimal employment set crucially depends on the relative concavity of the marginal information gain to the reservation utility.

In general, it can be shown that if the production function is a function of the sum of the types of employees, and it is submodular in set inclusion operation, then the production would exhibit decreasing marginal returns as shown in Gul and Stacchetti (1999). Then the optimal employment policy depends on the relative concavity of the marginal gain to the reservation utility, or quasi-convexity (quasi-concavity) of the marginal net gain. We raise two issues: what the characteristics of the optimal employment set are, and whether we can find a greedy algorithm in which local search leads to the global optimum. We leave these topics for future research.

Chapter 3

Uncertain Arrival Timing of a Signal and Forecasting Behavior

3.1 Introduction

Consider a forecaster who wishes to predict a firm's earnings. He has heard news from an insider which is informative for the forecasting. He also observes other forecasters' forecasts from which she may get some information on the earnings. He should now figure out whether the news has already been incorporated in the previous forecasts, since otherwise he may doubly count the same information.

This paper presents a parsimonious forecasting model to examine the effect of a common signal on the forecast when the arrival time of the signal is uncertain. We are in particular interested in the Bayesian updating procedure through which the forecaster fully utilizes the information contained in the consensus from past forecasters and the new signal. We then examine the empirical implication of the model.

Most literature on this topic, both in empirical and in theoretical studies, assumes a simple information structure in which every private signal is long-lived, and all public information is contained in the prior. On the contrary, some information in real world is short-lived, and shared by forecasters.[1]

This chapter examines another type of information. A signal is issued at some

[1] Even in the case where private signals are short-lived, if successors know who among predecessors have the signal and when it gets revealed, the outcome is equivalent to the assumption of a long-lived private signal.

point in time, but forecasters who receive it do not know when it is issued. Examples include leaked information from insiders, informative rumors, and public information transmitted slowly. In the model we present, the consensus may already reflect the new signal. Facing this uncertainty, the forecaster needs to estimate the chance that his information has been stale. This estimation procedure affects the estimate of the true state.

Under a Gaussian specification, it is shown that the forecaster's weight on his own information increases in the gap between his own information (the new signal) and the consensus. This is because, as the private signal of the forecaster moves out from the consensus, he believes his signal is more likely to be new. As a result, the forecast is more dispersed compared to the case with certain arrival timing.

The result sheds light on recent empirical studies on herding/anti-herding bias in earnings' forecasts. Chen and Jiang (2006) and Bernhardt et al. (2006) show that forecasters tend to anti-herd in the sense that they assign more weight on the private information than on the public information, or consensus, compared to rational Bayesian updating. Since the private information is not observable, they use the fact that the expected forecast error is uncorrelated with deviation from the consensus. In other words, the forecast error should not be predictable by available information.

Without resorting to behavioral assumption or strategic behavior due to certain features of the payoff, this paper shows that statisticians observe forecasters to weight more on private information rather than the consensus. This is contrary to the argument in Bernhardt et al. (2006) that their result is not affected by variations in the information structure.

The article is organized as follows. We present our model in the next section. We then analyze the Bayesian updating procedure for the forecaster and derive the properties of the forecasts. The empirical implications are proposed in the following section. In the last section, we discuss a few related issues and conclude.

3.2 Model

A forecaster tries to give her best forecast on the true state x. We call her *the forecaster* to distinguish her from past forecasters. A priori the true state is assumed to follow normal distribution with mean μ_x and precision τ_x, i.e., $x \sim \mathcal{N}(\mu_x, 1/\tau_x)$. We assume this prior mean μ_x is *unknown* to the forecaster, but τ_x is known. This is critical assumption for the main model. Instead, she observes a consensus from past forecasters. We denote the consensus by c.[2] The distribution of the consensus will be discussed later.

The forecaster observes an informative signal s and it is the sum of the true state and the error term where the error term is independent to the true state.[3] Formally, we have

$$s = x + \epsilon_s, \quad \epsilon_s \sim \mathcal{N}(0, 1/\tau_s)$$

The consensus can be interpreted as the aggregate information from past forecasters. In the perspective of the forecaster, the consensus c is the prior information available. The consensus provides information on true state, but the forecaster does not know whether it already contains the information from the signal she has observed. This is the case when the signal has arrived before the consensus is made, or before the current forecaster's turn.

To model this environment, we assume there are two events about the arrival time of the new signal, which is represented by a random variable $t \in \{0, 1\}$. If $t = 0$, the signal is arrived in the past and the signal is incorporated in the consensus. If $t = 1$, it is new to the forecaster. The ex-ante probability of the former event is $1 - p$ for $p \in (0, 1)$.

The observed consensus c now depends on whether the signal s has been arrived

[2]We slightly abuse notations to denote by c both the random variable and a realization of it.

[3]In principle, we need to assume the forecaster receives her own pure private information. We ignore this private information to focus only on the effect of common signal.

before the forecaster or he is the first forecaster who observes it. When $t = 1$, it is the mean of the distribution

$$\mathcal{N}\left(\mu_x, \frac{1}{\tau_x}\right),$$

while when $t = 0$, it is the mean of the distribution

$$\mathcal{N}\left(\frac{\tau_x \mu_x + \tau_s s}{\tau_x + \tau_s}, \frac{1}{\tau_x + \tau_s}\right)$$

Note that μ_x is unknown, which implies that observing s does not guarantee the forecaster knows the arrival timing of the signal.

After observing c and s, the forecaster should submit her forecast on the true state. It is assumed the compensation is decreasing in the ex-post mean squared error and it is paid when the true state is realized. Thus, it is optimal for the forecaster to submit the posterior mean of the true state given all information available to him. We now consider the optimal forecast by a Bayesian rational forecaster.

3.3 Bayesian Update and Optimal Forecast

As a benchmark, suppose first that there is no uncertainty in t, i.e., the forecaster knows which event with respect to t has occurred. Then, if $t = 0$, the forecaster has no additional information, and therefore her forecast would be

$$r = r_0 \equiv c,$$

while if $t = 1$, since she knows the signal s should be used to update the forecast of the true state, we have

$$r = r_1 \equiv \frac{\tau_x c + \tau_s s}{\tau_x + \tau_s}$$

Now, suppose that the event t is uncertain. Then, the forecaster needs to determine from which distribution the signal has been drawn. Let π be the posterior belief that $t = 1$. Then, for a K which depends on s and the belief on μ_x, we have

$$\begin{aligned} \frac{\pi}{1-\pi} &= \frac{Pr\left[t=1 \mid c, s\right]}{Pr\left[t=0 \mid c, s\right]} \\ &= \frac{f\left(c \mid t=1, s\right) Pr\left[t=1\right]}{f\left(c \mid t=0, s\right) Pr\left[t=0\right]} \\ &= K(s, \mu_x)\frac{p}{1-p} Exp\left[\frac{\tau_s}{2}(s-c)^2\right] \end{aligned}$$

This results in the following lemma.

LEMMA 3.1: *Keeping K constant, the posterior belief on $t = 1$ is increasing in $\mid s - c \mid$.*

PROOF. The likelihood ratio increases in $\mid s - c \mid$, and so does the probability. ■

The result is intuitive. The forecaster knows that the consensus is the mean of the true state conditional on the information past forecasters possess, which is the weighted mean of the prior mean and the signals they receive. As we can see in Figure 3.1, if the signal s is observed by past forecasters, the consensus must be close to the signal compared to the case without it regardless of the belief on μ_x. In other words, the closer c is to s, it is more likely that the consensus contains information of the common signal, or $t = 0$ has occurred. We now state the main result in the following proposition.

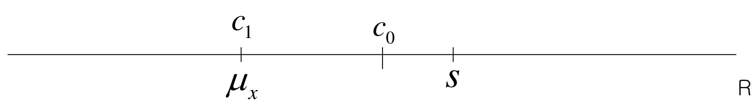

Figure 3.1: The consensus and a signal

PROPOSITION 3.1: *For a fixed s, with uncertain timing, the forecaster weighs more on the private signal as the signal moves away from the consensus.*

Proof. Note that

$$r = E[x \mid s, c] = Pr[t = 0 \mid s, c]E[x \mid s, c, t = 0] + Pr[t = 1 \mid s, c]E[x \mid s, c, t = 1]$$
$$= \pi r_1 + (1 - \pi)r_0$$

This implies that as s deviates farther from the consensus, the forecaster weight more on r_1, resulting in more weight on s. ■

3.4 Forecasting Error and Anti-Herding

In recent empirical studies, Zitzewitz, Eric (2001), Bernhardt et al. (2006), and Chen and Jiang (2006) argue that earnings' forecasters exhibit anti-herding, rather than herding that the previous literature had observed. They use the terminology anti-herding in the sense that forecasters weight more on private information than on the public information, or consensus, compared to the rational Bayesian updating. The possible explanations on the anti-herding in the literature are i) behavioral forecasting such as *overconfident* behavior, and ii) some payoff distortions which drive forecasters to forecast more riskily. In this section, we argue that it is explained by the uncertain arrival of some signals.

Bernhardt et al. (2006) base their argument on the fact that regardless of the distribution of information, as long as the forecast is unbiased in the sense that forecasters report their posterior mean or median, the probability that the realized earning is bigger than the report should be half. Similarly, Chen and Jiang (2006) examine the weight forecasters assign to private signal in an indirect way, and conclude that they assign more weight to private signal. In the following, we restate the main empirical model in Chen and Jiang (2006) in terms of our model, and examine the relationship between the forecasting error and the deviation from the consensus.[4]

[4] The probabilistic empirical model in Bernhardt et al. (2006) can be analyzed similarly. They use the fact that when the state turns out to be above the prior mean, the honest forecast tends to be lower than the realized value. In other words,

Since the public and private information are not observable, Chen and Jiang (2006) use the fact that, with the honest forecast on the true state, the expected forecast error is uncorrelated with deviation from the consensus. In other words, the forecasting error should not be predictable by available information.[5] This logic works for our model with certain timing. To see this, fix $t = 1$ and consider s as a private signal. Let h be the weight on s in Bayesian update, k be the actual weight on the signal in the actual forecast, and r be the actual forecast. In our model, $h = \tau_s/(\tau_x + \tau_s)$. Then, if the forecast is the posterior mean of the true state,

$$E[x \mid c, s] = hs + (1 - h)c.$$

But since $r = ks + (1 - k)c$,

$$E[FE \mid c, s] = E[r - x \mid c, s] = \frac{k - h}{k}(r - c) = \beta_0 Dev.$$

for a constant β_0, or $k = h$. If the actual report is the posterior mean of the true state, it must be that $\beta_0 = 0$.

In our model, we assume the signal s could be either private or public. Nonetheless, if the timing is certain, the expected forecast error is uncorrelated to the deviation since without the private signal ($t = 0$) the deviation is zero. We now consider the case of uncertain timing. The following proposition shows if the signal deviates from the consensus far enough, then the expected forecast error is positively correlated with the deviation.

$$E\left[r_i \mid x > \mu_x\right] = E\left[\frac{\mu_x \tau_x + s_i \tau_i}{\tau_x + \tau_i} \mid x > \mu_x\right] = \frac{\mu_x \tau_x + x\tau_i}{\tau_x + \tau_i} < x.$$

However, this analysis is not valid if the prior mean is stochastic as is our model.

[5]Indeed, the expected forecasting error is uncorrelated with the forecast. In general, for the optimal forecast r_i by the forecaster with private signal s_i, we have

$$E\left[r_i(s_i)E_x(r_i(s_i) - x)\right] = E\left[E(x \mid s_i)E\left(E(x \mid s_i) - x\right)\right] = E\left[E(x \mid s_i)E\left(E(x \mid s_i) - x \mid s_i\right)\right] = 0$$

PROPOSITION 3.2: *Suppose $\pi > p$ ($\pi < p$). Then, the expected forecast error given c and s is positively (negatively) correlated with the deviation from the consensus.*

PROOF. The following calculations are straightforward.

$$E\left[r - x|c, s, t = 1\right] = \pi\frac{\tau_x c + \tau_s s}{\tau_x + \tau_s} + (1-\pi)c - \frac{\tau_x c + \tau_s s}{\tau_x + \tau_s} = (1-\pi)\frac{\tau_s(c-s)}{\tau_x + \tau_s},$$

$$E\left[r - x|c, s, t = 0\right] = \pi\frac{\tau_x c + \tau_s s}{\tau_x + \tau_s} + (1-\pi)c - c = \pi\frac{\tau_s(s-c)}{\tau_x + \tau_s}.$$

The forecast error $FE \equiv r - x$ is now in expectation,

$$\begin{aligned} E\left[FE \mid c, s\right] &= pE\left[r - x|c, s, t = 1\right] + (1-p)E\left[r - x|c, s, t = 0\right] \\ &= p(1-\pi)\frac{\tau_s(c-s)}{\tau_x + \tau_s} + (1-p)\pi\frac{\tau_s(s-c)}{\tau_x + \tau_s} \\ &= \frac{\tau_s(\pi - p)}{\tau_x + \tau_s}(s - c) \qquad (3.1) \end{aligned}$$

Recall that the forecast r with uncertain timing is

$$r = \pi\frac{\tau_x c + \tau_s s}{\tau_x + \tau_s} + (1-\pi)c.$$

Therefore, the deviation of the forecast from the consensus, $Dev \equiv r - c$, follows

$$r - c = \pi\frac{\pi\tau_s}{\tau_x + \tau_s}(s - c).$$

From (3.1),

$$E\left[FE \mid c, s\right] = \frac{\pi - p}{\pi}(r - c),$$

which completes the proof. ■

Proposition 3.2 shows that when the signal timing is uncertain, the forecast error is not independent with the deviation from the consensus. The sign of the correlation depends on the the relationship between the ex-ante and ex-post belief on the arrival timing. Note that in the previous empirical literature, the positive correlation is considered as an evidence of anti-herding. In our model, the statistician may observe either herding or anti-herding, depending on the realization of the consensus and signal.

The result may be extended to the expected forecasting behavior, considering the distribution of the signal and the belief on the prior mean. This is left for the future research, but we can propose some conjecture here. Sine the ex-post belief increases in the gap between the consensus and the signal, as the signal distribution has higher variance or the consensus is less informative, it is more likely for the statistician to observe anti-herding.

3.5 Conclusion and Discussion

In this chapter, we examine through a parsimonious model the effect of common information on the forecast when the arrival timing of the information is uncertain. We show that the posterior distribution of the arrival timing depends on the deviation of the signal from the consensus. This implies that the optimal forecast also depends on the deviation, which may result in the seemingly anti-herding behavior. The result has an empirical implication: without resorting to the behavioral or payoff-relevant distortion, we can observe that the expected forecast error may be positively correlated with the deviation.

Though we explain the puzzle raised in recent empirical studies on anti-herding behavior in earnings' forecasts through a non-standard information structure, there are other models consistent with the finding such as one with a behavioral approach and one with non-standard payoff. Which model provides a more plausible explanation would be an important and interesting empirical issue.

The generalization of the model is left for future research. The final goal of the generalization would be explaining the time series properties of actions (buy/sell decision, forecast, etc.) when public information arrives with stochastic timing. The main issue here is how the learning procedure affects the assessment of the timing, and we believe that it would provide some empirical regularities such as the short-term under-reaction and the long-term over-reaction to news event.[6]

[6]Short-term momentum and long-term mean reversion is a version of this regularity in the financial market. As far as we know, the only plausible model to explain this is through the behavioral approach, for example in Hong and Stein (1999).

Bibliography

[1] Abrue, Dilip and Markus Brunnermeier (2002a), "Synchronization Risk and Delayed Arbitrage," *Journal of Financial Economics*, Vol. 66.

[2] Abrue, Dilip and Markus Brunnermeier (2002b), "Bubble and Crashes," *Econometrica*, Vol. 71, No. 1.

[3] Avery, C. N. and J. A. Chevalier (1999), "Herding over the Career." *Economic Letters*, Vol. 63.

[4] Bernhardt, Dan, Murillo Campello, and Edward Kutsoati(2006), "Who Herds?" *Journal of Financial Economics*, Vol. 80, No. 3.

[5] Bhattacharya, Sudipto and Paul Pfleiderer (1985), "Delegated Portfolio Management." *Journal of Economic Theory*, Vol. 36, No. 1.

[6] Brunnermeier, Markus (2005), "Information Leakage and Market Efficiency," *Review of Financial Studies*, Vol. 18, No. 2.

[7] Chade, Hector and Lones Smith (2005), "Simultaneous Search," Mimeo.

[8] Chade, Hector and Lones Smith (2006), "Simultaneous Search," *Econometrica*, Vol. 74, No.5.

[9] Chen, Qi and Wei Jiang(2006), "Analysts' Weighting of Private and Public Information," *The Review of Financial Studies*, vol. 19, no. 1.

[10] Chu, Jennifer(2005), Does Bayesian Updating Occur? Evidence from Analyst Earnings Forecasts, mimeo

[11] Crawford, V. and Joel Sobel (1982), "Strategic Information Transmission." *Ecoometrica*, Vol. 50, No. 6.

[12] Crémer, Jacques and Richard P. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, Vol. 53, No. 2.

[13] Crémer, Jacques and Richard P. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, Vol. 56, No. 6.

[14] Fang, Fang, Maxwell Stinchcombe, and Andrew Whinston (2007), ""Putting Your Money Where Your Mouth Is" - A Betting Platform for Better Prediction," *Review of Network Economics*, Vol. 6, Issue 2.

[15] Green, J. and N. Stokey (1983), "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, Vol. 91, No. 3.

[16] Gul, Faruk and Ennio Stacchetti (1999), "Walrasian Equilibrium with Gross Substitutes," *Journal of Economic Theory*, Vol. 87, 95-124.

[17] Hansen, Robert G. (1985) "Auctions with Contingent Payments," *The American Economic Review*, Vol. 75, No. 4.

[18] Hatfield, William and Paul R. Milgrom (2005), "Matching with Contracts," *American Economic Review*, Vol. 95, No. 4.

[19] Holmstrom, B. (1982), "Moral Hazard in Teams," *The Bell Journal of Economics*, Vol. 13, No. 2.

[20] Hong, Harison and Jeremy C. Stein (1999), "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *The Journal of Finance*, Vol. LIV., No. 6.

[21] Jullien, Bruno (2000), "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, Vol. 92, No. 1.

[22] Kelso, A. S. Jr., and V. P. Crawford (1982), "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, Vol. 50, No. 6.

[23] Lehmann, Benny, Daniel Lehmann, and Norman Nissan (2006), "Combinatorial auctions with decreasing marginal utilities," *Games and Economic Behavior*, Vol. 55, 270-296.

[24] Levy, G. (2004), "Anti-Herding and Strategic Consultation." *European Economic Review*, Vol. 84.

[25] Lovàsz, L. (1982), "Submodular Functions and Convexity," *Mathematical Programming: The State of the Art*, ed. by A. Bachem, M. Grotschel, and B. Korte, Springer, Berlin.

[26] Maskin, Eric and John Riley (1984), "Monopoly with Incomplete Information." *Rand Journal of Economics*, Vol. 15, No. 2.

[27] Markov, Stanimir and Ane Tamayo(2003), "Predictability in Financial Analyst Forecast Errors: Learning or Irrationality?" mimeo.

[28] Milgrom, Paul, and Bruno Strulovici (1996), "Substitute Valuation, Auctions, and Equilibrium with Discrete Goods," *mimeo.*

[29] Murota, Kazuo and Akiyoshi Shiourab (2003), "Quasi M-convex and L-convex functions - quasiconvexity in discrete optimization," *Discrete Applied Mathematics*, Vol. 131, pp. 467-494.

[30] Osband, Kent (1989), "Optimal Forecasting Incentives" *Journal of Political Economy*, Vol. 97, No. 5.

[31] Ottaviani, Marco and Peter N. Sørensen (2006a), "Reputational Cheap Talk" *RAND Journal of Economics*, Vol. 37, No. 1.

[32] Ottaviani, Marco and Peter N. Sørensen (2006b), "The Strategy of Professional Forecasting," *Journal of Financial Economics*, Vol. 81, No. 2.

[33] Prendergast, Canice (1993), "A Theory of "Yes Man"," *The American Economic Review*, Vol. 83, No. 4..

[34] Pycia, Marek (2007), "Many-to-One Matching with Complementarities and Peer Effects," mimeo.

[35] Scharfstein, D. and J. Stein (1990), "Herd Behavior and Investment." *American Economic Review*, Vol. 80, No. 3.

[36] Sorenson, Eric H. (1985), "In Defense of Technical Analysis: Discussion," *The journal of Finance*, Vol. 80, No. 3.

[37] Treynor, Jack L., and Robert Fugerson (1985) "In Defense of Technical Analysis," *The Journal of Finance*, Vol. 40, No. 3.

[38] Trueman, B. (1994), "Analyst Forecasts and Herding Behavior." *Review of Financial Studies*, Vol. 7, No. 1.

[39] Zitzewitz, Eric (2001), "Measuring Herding and Exaggeration by Equity Analysts and Other Opinion Sellers," Stanford University Working Paper.