Optimal contracts and career concerns: Comparing information structures*

Michiko Ogaku

Graduate School of Systems and Information Engineering, University of Tsukuba,

1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan; ohgaku@sk.tsukuba.ac.jp

4 August 2014

Abstract

This paper models optimal contracts in the presence of career concerns in a general distribution formulation in which information structures are comprised of probability density functions of contractible and non-contractible signals. The model provides a general threshold level of information structure where the optimal contract is a fixed payment or a penalty contract. Furthermore, this paper provides properties of information structures in a special case that non-contractible signals are sufficient statistics.

Keywords: Optimal contracts, Career concerns, Information structures.

1 Introduction

This paper adds contractible signals to Dewatripont et al. (1999) in order to investigate the impact of career concerns on the optimal contracts in a general distribution formulation. The model provides a general threshold level of information structure where the

*This paper is a revised version of Discussion paper #1307 of Department of Social Systems and Management, University of Tsukuba.

optimal contract is a fixed payment or violates the condition for it to be non-decreasing, that is, a higher contractible signal realisation may not always imply a higher compensation. The threshold tells that even when optimal contracts are monotone non-decreasing in the contractible signal in a standard principal-agent model, it can be a fixed payment or a penalty contract in the presence of career concerns. In addition, this paper finds that if contractible signals are noninformative in the sense of Holmström (1979) (in other words, non-contractible signals are sufficient statistics), then the optimal contract is a fixed payment or a penalty contract. Furthermore, this paper shows that how the finding relates to the MPS criterion of Kim (1995) which plays a prominent role in information system rankings for a standard principal-agent relationship. Finally, this paper provides how garbling of information in the sense of Blackwell (1951) affects optimal contracts and implicit incentives.

In this general distribution formulation, the conditional probability density function of the contractible signal given the observed non-contractible signal does not always hold the MLRP (monotone likelihood ratio property) — the property that higher signals are good news, even though the contractible signal individually holds the property and it tells whether the optimal incentive scheme on the basis of the contractible signal can be monotone non-decreasing.

The central inequation in this paper (inequation (8), the threshold mentioned above) is expressed in a comparison of covariances of each signal's likelihood ratio and the agent's output. The expression allows to solve for the relationship with Kim's MPS criterion. Considering as if the non-contractible signal is a candidate for observables used for designing an incentive scheme, if the information system of non-contractible signal is more efficient than that of contractible signal in the sense of Kim's MPS criterion, then the optimal incentive scheme cannot be monotone non-decreasing. This condition is sufficient but not necessary for the central inequation of this paper in (8). It implies that an efficient information system in a standard principal-agent could violate conditions for the incentive scheme to be progressive in this model.

Now how this paper relates to prior work is explained. This paper generalises Gibbons and Murphy (1992) to a general distribution formulation. Gibbons and Murphy (1992) first show how incentive contracts are influenced by career concerns using the Normal-Exponential (CARA-Gaussian) model restricting linear contracts. The CARA Gaussian model considering both incentive contracts and career concerns are utilised in Meyer and Vickers (1997), Kaarbøe and Olsen (2008), Sabac (2007, 2008) and Ogaku (2014). However, prior work remains silent about any property of information structures where the interplay between contracts and career concerns is considered. With regard to implicit incentives from career concerns, Dewatripont et al. (1999) shows that information structures can be compared in an analogous manner to information system in a standard principal-agent model. The present paper starts by mixing career concerns in a general form, principal-agent problems and the first order approach to solving the problems.

The subject of information system rankings was first addressed by Blackwell (1951). It was introduced as a decision maker's (statistician) ranking of experiments (sampling procedures). Gjesdal (1982) analysed it as a problem of information system choice in a principal-agent problem. He showed that the Blackwell's ranking is sufficient but not necessary for many principal-agent problems.[1] While Holmström (1979) independently proposed the informativeness criterion which plays a leading part in information system rankings for a principal-agent problem. In essence, the informativeness criterion means that additional information should be used for designing contracts until sufficient statistics are obtained because at this point, further information does not add any news that will affect the agent's action. More recently, introducing his MPS criterion, Kim (1995) extended Holmstöm's informativeness criterion which supposes inclusive information systems[2] by making it also available to rank noninclusive information systems.[3] This paper offers new insights into the literature by showing that non-contractible information might invalidate information system rankings on the basis of prior work.

Sufficient conditions for the optimal contract to be non-decreasing have intensively

[1] Gjesdal (1982) showed that the Blackwell's ranking might be invalid when the agent's risk preferences depend on his action.

[2] Inclusive information systems can be described as a vector x and a subvector $T(x)$ that is obtained by deleting some components of x.

[3] Kim (1995) proved that Holmström's informativeness criterion and the MPS criterion are equivalent when the ranking is conducted between inclusive information systems.

been investigated in a standard principal-agent paradigm (Grossman and Hart, 1983; Rogerson, 1985; Sinclair-Desgagné, 1994). Grossman and Hart (1983) and Rogerson (1985) showed that if MLRP and CDFC (concavity of distribution function condition) hold, then a second-best optimal incentive scheme satisfies the monotonicity constraint when the signal space is one dimension. With regard to that in a multi-dimensional signal space, Sinclair-Desgagné (1994) confirmed that MLRP and generalised CDFE in multi-dimensional space are sufficient conditions. This paper assumes MLRP but not CDFC, however, what makes the monotonicity invalid is non-contractible information's relative efficiency in developing a reputation in the market.

The basic model is in Section 2. Section 3 presents the threshold for optimal contracts violating non-decreasing monotonicity. Section 4 shows how the central inequation in (8) relates to Kim's MPS criterion. Section 5 shows how the garbling signals in the sense of Blackwell's theorem impacts on optimal contracts and implicit incentives. Section 6 concludes.

2 Basic Model

Consider a single-period principal-agent relationship in the competitive labour market, where a risk-averse agent with unknown talent $\theta \in \mathbb{R}$ privately takes an action $a \in A \subset \mathbb{R}$. The agent's talent and action will be learned by a set of random vectors $\boldsymbol{X} = (X_1, \ldots, X_m)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ which becomes commonly observable without cost. This paper assumes that $\boldsymbol{Y}$ is contractible but $\boldsymbol{X}$ is not contractible. Profits of contracts and optimal contracts are defined as a function of contractible signals, $\pi : \mathbb{R}^n \to \mathbb{R}$ and $s : \mathbb{R}^n \to \mathbb{R}$, respectively, however, as is well known in the career concern literature, non-contractible signals can work as an incentive device in the form of the agent's concerns about his reputation in the market. Let the reputation be represented by the market's equilibrium expectation of the agent's talent after he/she observes the full statistic $(\boldsymbol{X}, \boldsymbol{Y})$. Let $a^* \in A$ be an equilibrium action, then the reputation is written as $R(\boldsymbol{x}, \boldsymbol{y}) = E[\theta | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}, a^*]$.

The information structure is represented by a probability density function of $(\boldsymbol{X}, \boldsymbol{Y}, \theta)$ parametrised by the agent's action, $f(\boldsymbol{x}, \boldsymbol{y}, \theta|a)$.

The analysis begins by imposing the following assumptions.

Assumption 1: The agent has the following utility function:

$$U(w) - V(a), \quad U' > 0, \quad U'' < 0, \quad V' > 0, \quad V'' > 0,$$

where $w = s + R$ is the aggregation of the contract s and the reputation R, and V denotes a measure of the agent's disutility of effort.

Assumption 2: $f(\theta, \boldsymbol{x}, \boldsymbol{y}|a)$ is positive and twice continuously differentiable and respective marginal densities are

$$\begin{aligned} \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) &= \int f(\boldsymbol{x}, \boldsymbol{y}, \theta|a) d\theta, \\ \hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}|a) &= \int \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x}, \ \textit{and} \\ \hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a) &= \int \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{y}, \end{aligned}$$

where $\hat{f}_{\boldsymbol{Y}}$ has the MLRP.

Assumption 3: The reputation $R(\boldsymbol{x}, \boldsymbol{y})$ is additively separable in $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$R(\boldsymbol{x}, \boldsymbol{y}) = A(\boldsymbol{y}) + B(\boldsymbol{x}),$$

and $(\boldsymbol{Y}, \boldsymbol{X}, \theta)$ is conditionally increasing sequence (CIS).[4] It is well known (see e.g. Müller and Stoyan, 2002, chap. 3) that if $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ is CIS then $\boldsymbol{Z}$ is associated.[5]

Let $\tilde{s}(\boldsymbol{y}) = s(\boldsymbol{y}) + A(\boldsymbol{y})$. $\tilde{s}(\boldsymbol{y})$ is the agent's overall reward from the contractible variable $\boldsymbol{y}$ and is used in the prospective employers' problem explained below. Note that

[4] A random variable $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ is CIS if $E[\phi(Z_i)|Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}]$ is an increasing function of the variables $z_1, \ldots, z_{i-1}$ for all increasing function ϕ for which the expectations are defined, $i = 2, \ldots, d$ (see e.g. Müller and Scarsini, 2001). One of examples of random variables that satisfy Assumption 3 is given in Example 1.

[5] A random variable $\boldsymbol{Z}$ on a partially ordered Polish space S is associated if $Cov(g(\boldsymbol{Z}), h(\boldsymbol{Z})) \geq 0$ for all increasing $f, g : S \to \mathbb{R}$.

the prospective employers cannot control the non-contractible signal $\boldsymbol{x}$ by definition.

Assumption 4: $\pi(\boldsymbol{y})$ is non-decreasing in $\boldsymbol{y}$.

In the competitive labour market the prospective employers are supposed to be wage takers, i.e. they compete, resulting in the optimal reward $\tilde{s}(\boldsymbol{y})$ maximising the agent's expected utility. The optimal $\tilde{s}$ is a solution of the problem:

$$\max_{\tilde{s},a} \iint \left(U(w) - V(a)\right) \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y} \tag{1}$$

subject to

$$\iint \Big(\pi(\boldsymbol{y}) + \int \theta \frac{f(\boldsymbol{x}, \boldsymbol{y}, \theta|a^*)}{\hat{f}(\boldsymbol{x}, \boldsymbol{y}|a^*)} d\theta - w\Big) \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y} = 0 \quad \textit{and} \tag{2}$$

and

$$\iint U(w) \hat{f}_a d\boldsymbol{x} d\boldsymbol{y} = V'(a), \tag{3}$$

where (2) reflects the restriction that the contract must earn zero expected profit for the principal. In other words, it is the participation constraint for the agent; the principal must offer the agent the expected reward w at least as high as the agent's expected outcome $\pi(\boldsymbol{y})$ and talent $R(\boldsymbol{x}, \boldsymbol{y})$. (3) represents the incentive constraint. Let the agent's expected utility $M(a) \equiv \iint \left(U(w) - V(a)\right) \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y}$. (3) is a relaxed constraint of $a \in \arg\max_{\hat{a} \in A} M(\hat{a})$ and this approach is called the first-order approach.[6] (3) represents a stationary point for the agent, i.e. $M'(a) = 0$. In order to guarantee that the point be the agent's optimal action choice, $M(a)$ is supposed to be strictly concave, i.e. $M''(a) < 0$ for all $a \in A$. Consider the Lagrangian form obtained by assigning undetermined multipliers

[6] Grossman and Hart (1983), Rogerson (1985) and Sinclair-Desgagné (1994) showed that MLRP and CDFC are sufficient for the validity of the first-order approach.

λ and μ to (2) and (3):

$$
\begin{aligned}
L(\tilde{s}, a) = &\iint \left(U(w) - V(a)\right) \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y} \\
&+ \lambda \iint \Big(\pi(\boldsymbol{y}) + \int \theta \frac{f(\boldsymbol{x}, \boldsymbol{y}, \theta|a^*)}{\hat{f}(\boldsymbol{x}, \boldsymbol{y}|a^*)} d\theta - w\Big) \hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y} \\
&+ \mu \Big(\iint U(w) \hat{f}_a d\boldsymbol{x} d\boldsymbol{y} - V'(a)\Big).
\end{aligned}
$$

Differentiating with respect to the scalar a and the function $\tilde{s}$,[7] one obtains the first-order conditions:

$$
\frac{\partial L}{\partial a} = M'(a) + \lambda \iint \Big(\pi(\boldsymbol{y}) + \int \theta \frac{f(\boldsymbol{x}, \boldsymbol{y}, \theta|a^*)}{\hat{f}(\boldsymbol{x}, \boldsymbol{y}|a^*)} d\theta - w\Big) \hat{f}_a(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y} + \mu M''(a) = 0,
$$

i.e.

$$
\mu = -\lambda M''(a)^{-1} \iint \Big(\pi(\boldsymbol{y}) + \int \theta \frac{f(\boldsymbol{x}, \boldsymbol{y}, \theta|a^*)}{\hat{f}(\boldsymbol{x}, \boldsymbol{y}|a^*)} d\theta - w\Big) \hat{f}_a(\boldsymbol{x}, \boldsymbol{y}|a) d\boldsymbol{x} d\boldsymbol{y}, \tag{4}
$$

and

$$
\frac{\partial L}{\partial \tilde{s}} = U'(w)\hat{f} - \lambda \hat{f} + \mu U'(w) \hat{f}_a = 0,
$$

i.e.

$$
\frac{\lambda}{U'(w)} = \left(1 + \mu \frac{\hat{f}_a}{\hat{f}}\right). \tag{5}
$$

Substituting the equilibrium effort $a = a^*$, μ in (4) can be written as

$$
\mu = -\lambda M''(a^*)^{-1} cov\Big(\pi(\boldsymbol{y}) - s(\boldsymbol{y}), \frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}\Big). \tag{6}
$$

The expression $\pi(\boldsymbol{y}) - s(\boldsymbol{y})$ on left-hand side of (6) denotes the share that goes to the principal. If $\pi(\boldsymbol{y}) - s(\boldsymbol{y})$ is non-decreasing in $\boldsymbol{y}$, then (6) is consistent with Proposition 1 that will be explained in Section 3.

[7] A variation method is used with regard to w.

3 Monotonicity of optimal contracts

Let $s_f(\boldsymbol{y})$ be the optimal incentive scheme that depends on the information structure f and satisfies (4) and (5). Correspondingly, let $\tilde{s}_f(\boldsymbol{y})$ be the agent's overall reward from $\boldsymbol{y}$ when $s_f(\boldsymbol{y})$ is optimal.

Observation 1. *If $\mu > 0$ and $\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_a}{\hat{f}}\right) \geq 0$, then $\tilde{s}'_f(\boldsymbol{y}) \geq 0$ and the equation is satisfied only if $\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_a}{\hat{f}}\right) = 0$ is satisfied.*

Proof. Differentiating (5) with respect to $\boldsymbol{y}$, one obtains the following.

$$-\lambda \frac{U''}{(U')^2} w_{\boldsymbol{y}} = \mu \frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_a}{\hat{f}}\right). \tag{7}$$

Let the first-best reward when the agent's incentive constraint in (3) is not considered be w_λ. w_λ must satisfy $\lambda = U'(w_\lambda)$. Since $U' > 0$ from Assumption 1, $\lambda > 0$ must be satisfied. With Assumption 1, Assumption 2 and $\lambda > 0$, (7) implies $w_{\boldsymbol{y}} = \tilde{s}'_f(\boldsymbol{y}) \geq 0$ and the equation is satisfied only if $\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_a}{\hat{f}}\right) = 0$ is satisfied. □

Proposition 1. *Suppose Assumption 1 and 3 hold. Then, any μ satisfying* (3) *and* (5) *is positive.*

The proof is in Appendix.

Observation 1 and Proposition 1 immediately provide the following corollary.

Corollary 1. *$s_f(\boldsymbol{y})$ violates non-decreasing monotonicity or $s_f(\boldsymbol{y})$ is fixed payment if the contractible signal $\boldsymbol{y}$ is noninformative in the sense of Holmström (1979), i.e. there exists function $g : \mathbb{R}^{m+n} \to \mathbb{R}$ such that for all $(\boldsymbol{x}, \boldsymbol{y}, a)$, the density $\hat{f}$ can be factorised according to*

$$\hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) = g(\boldsymbol{x}, \boldsymbol{y})\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a).$$

Proof. The equality

$$\hat{f}(\boldsymbol{x}, \boldsymbol{y}|a) = g(\boldsymbol{x}, \boldsymbol{y})\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a)$$

implies

$$\frac{\hat{f}_a}{\hat{f}} = \frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}.$$

Thus, one has

$$\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_a}{\hat{f}}\right) = 0.$$

From the above equation, Observation 1 and Proposition 1, one has $\tilde{s}'_f(\boldsymbol{y}) = 0$. From Assumption 3, $s_f(\boldsymbol{y})' \leq 0$ is implied. □

The proof of Corollary 1 demonstrates that an additional signal $\boldsymbol{y}$ is informative as long as it conveys information about an optimal reward/penalty on the basis of $\boldsymbol{y}$. Thus, the converse of Corollary 1 is false.

Corollary 1 is straightforward; when contractible signals are uninformative, their information systems do not hold a condition in which optimal contracts are non-decreasing. The following proposition provides the threshold that $s_f(\boldsymbol{y})$ fails to hold non-decreasing monotonicity.

Proposition 2. *$s_f(\boldsymbol{y})$ violates non-decreasing monotonicity or $s_f(\boldsymbol{y})$ is fixed payment if and only if there exists $y_i, i = 1, \ldots, n$ such that*

$$cov\left(\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}, y_i\right) \geq cov\left(\frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}, y_i\right). \tag{8}$$

The proof is in Appendix.

The inequation (8) means that the conditional probability density function of the contractible signal $\boldsymbol{y}$ given the observed non-contractible signal $\boldsymbol{x}$ does not hold the MLRP, although the marginal probability density function $f_{\boldsymbol{Y}}(\boldsymbol{y})$, independently holds the property. In other words, it is when $\boldsymbol{y}$ is good news about the agent's effort but bad news if $\boldsymbol{x}$ is observed beforehand. If any $y_i, i = 1, \ldots, n$, holds (8), fixed payments are preferable.

Example 1. Let signals be

$$x = \theta + a + \nu, \quad x \sim N(\bar{\theta} + a, \sigma_\theta^2 + \sigma_\nu^2),$$
$$y = \theta + pa + \varepsilon, \quad y \sim N(\bar{\theta} + pa, \sigma_\theta^2 + \sigma_\varepsilon^2), \quad p \in (0, 1).$$

Then, covariances are given by

$$cov\left(\frac{\hat{f}_{Xa}}{\hat{f}_X}, y\right) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\nu^2},$$
$$cov\left(\frac{\hat{f}_{Ya}}{\hat{f}_Y}, y\right) = p,$$

and w_y is given by

$$\tilde{s}'_f(y) = \frac{cov\left(\frac{\hat{f}_{Ya}}{\hat{f}_Y}, y\right) - cov\left(\frac{\hat{f}_{Xa}}{\hat{f}_X}, y\right)}{Var(y|x, a)}.$$

The sign of $\tilde{s}'_f(y)$ can be positive and negative depending on the covariance relation of likelihood ratios. Given the fact that normal distributions do not obey CDFC, it implies that the monotonicity of optimal contracts in this model do not depend on CDFC.

The inequation in (8) in Example 1 is written as

$$p \leq \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\nu^2} \ (< 1),$$

where p denotes the marginal impact of the agent's action on y, σ_θ^2 is the variance of the agent's talent and σ_ν^2 is the variance of x's noise term. It implies that x is more responsive to the agent's action than y, i.e. $1 > p$. (8) holds when y is less responsive to the agent's action than x, both signal conveys opaque information about his talent but x's noise term is small.

4 Relationship with MPS Criterion

Now, this section shows how the inequation in (8) relates to Kim's MPS criterion. For completeness, this section restates Kim's result:

MPS Criterion: Assuming that the first-order approach is valid,[8] the information system $\hat{f}_{\boldsymbol{X}}$ is more efficient than $\hat{f}_{\boldsymbol{Y}}$ at $a = \hat{a}$ if the random variable $\frac{\hat{f}_{\boldsymbol{X}^a}(\boldsymbol{x}|\hat{a})}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|\hat{a})}$ is a mean preserving spread of $\frac{\hat{f}_{\boldsymbol{Y}^a}(\boldsymbol{y}|\hat{a})}{\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}|\hat{a})}$. That is,

$$\int^{z} L^{a}_{\hat{f}_{\boldsymbol{X}}}(t)dt \geq \int^{z} L^{a}_{\hat{f}_{\boldsymbol{Y}}}(t)dt \quad for\ all\ z \in \mathbb{R} \tag{9}$$

with the strict inequality holding for some range of $z \in \mathbb{R}$ with positive measure, where $L^{a}_{\hat{f}_{\boldsymbol{X}}}$ and $L^{a}_{\hat{f}_{\boldsymbol{Y}}}$ are the cumulative distribution function of $\frac{\hat{f}_{\boldsymbol{X}^a}(\boldsymbol{x}|\hat{a})}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|\hat{a})}$ and $\frac{\hat{f}_{\boldsymbol{Y}^a}(\boldsymbol{y}|\hat{a})}{\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}|\hat{a})}$, respectively.

It is well known that (9) implies that $\{\frac{\hat{f}_{\boldsymbol{X}^a}}{\hat{f}_{\boldsymbol{X}}}, \frac{\hat{f}_{\boldsymbol{Y}^a}}{\hat{f}_{\boldsymbol{Y}}}\}$ can be seen as a martingale which is useful to compare information structures. This section refers the result from Shaked and Shanthikumar (2007).

Lemma 1. (Shaked and Shanthikumar, 2007, Theorem 3.A.1 and Theorem 3.A.4).

$$(9)\ \ holds \Leftrightarrow \frac{\hat{f}_{\boldsymbol{Y}^a}}{\hat{f}_{\boldsymbol{Y}}} \leq_{cx} \frac{\hat{f}_{\boldsymbol{X}^a}}{\hat{f}_{\boldsymbol{X}}} \Rightarrow \quad \left\{\frac{\hat{f}_{\boldsymbol{Y}^a}}{\hat{f}_{\boldsymbol{Y}}}, \frac{\hat{f}_{\boldsymbol{X}^a}}{\hat{f}_{\boldsymbol{X}}}\right\} \quad is\ a\ martingale,$$

that is

$$\frac{\hat{f}_{\boldsymbol{Y}^a}}{\hat{f}_{\boldsymbol{Y}}} = E\left[\frac{\hat{f}_{\boldsymbol{X}^a}}{\hat{f}_{\boldsymbol{X}}}\middle|\boldsymbol{y}\right]. \tag{10}$$

Proof of Lemma 1 is found in Shaked and Shanthikumar (2007).[9]

Kim's MPS criterion is originally used for ranking information systems for optimal contracts but $\hat{f}_{\boldsymbol{X}}$ is the probability density function of the non-contractible signal $\boldsymbol{x}$, but following corollary temporally considers $\hat{f}_{\boldsymbol{X}}$ as if a candidate of information systems for

[8]This section interprets the assumption of the validity of the first-order approach as the validity in terms of $\hat{f}_{\boldsymbol{X}}$ and $\hat{f}_{\boldsymbol{Y}}$, separately.

[9]Dewatripont et al. (1999) and Shaked and Shanthikumar (2007) document that the MPS criterion is necessary and sufficient for the martingale property. This section uses only the sufficiency condition.

optimal contracts in order to compare $\hat{f}_{\boldsymbol{X}}$ and $\hat{f}_{\boldsymbol{Y}}$ on the basis of Kim's MPS criterion.

Corollary 2. *If information system $\hat{f}_{\boldsymbol{X}}$ is more efficient than $\hat{f}_{\boldsymbol{Y}}$ in the sense of Kim's MPS criterion, $s_f(\boldsymbol{y})$ violates non-decreasing monotonicity.*

Proof. For any functions $g : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned}
cov\left(\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}, g(\boldsymbol{y})\right) &= \iint g(\boldsymbol{y})\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\hat{f}(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} \\
&= \iint g(\boldsymbol{y})\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\hat{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} \\
&= \int g(\boldsymbol{y})\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y})\int \frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\hat{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} \\
&= cov\left(g(\boldsymbol{y}), E\left[\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\middle|\boldsymbol{y}\right]\right) \\
&= cov\left(g(\boldsymbol{y}), \frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}\right) \ (\because \text{ Lemma 1.}).
\end{aligned}$$

From Proposition 2, $s_f(\boldsymbol{y})$ violates non-decreasing monotonicity. □

Corollary 3. *The converse of Corollary 2 is false.*

Proof. Suppose there is $y_i, i = 1, \ldots, n$, such that $cov\left(y_i, \frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\right) > cov\left(y_i, \frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}\right)$. From the proof of Corollary 3, it implies information system $\hat{f}_{\boldsymbol{X}}$ is not more efficient than $\hat{f}_{\boldsymbol{Y}}$ in the sense of MPS criterion. □

In words, Corollary 2 and 3 means that Kim's MPS criterion is sufficient but not necessary for (8). It implies that how much $\hat{f}_{\boldsymbol{Y}}$ is efficient in a standard principal-agent framework, if there exists $\hat{f}_{\boldsymbol{X}}$ that is more efficient than $\hat{f}_{\boldsymbol{Y}}$, $\hat{f}_{\boldsymbol{Y}}$ can no longer be an information system which hold a primary condition in which optimal contracts are monotone non-decreasing. Moreover, even if $\hat{f}_{\boldsymbol{Y}}$ is efficient than $\hat{f}_{\boldsymbol{X}}$ in Kim's MPS criterion, $\hat{f}_{\boldsymbol{Y}}$ is not an information system which hold incentive scheme's non-decreasing monotonicity when (8) holds.

5 Garblings of information

This section considers a special case:

- contractible signals are garbling version of non-contractible signals in the sense of Blackwell (1951).

In the special case, optimal contracts and implicit incentives lead to some interesting applications.

By garbling of information, this section means applying a stochastic transformation on a signal to compose a new one. Assume for all $a \in A$ there exists a conditional density $\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ (a garbling) such that

$$\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}|a) = \int \hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a)d\boldsymbol{x}. \tag{11}$$

The information structure $\hat{f}_{\boldsymbol{Y}}$ in (11) is a garbling version of the information structure $\hat{f}_{\boldsymbol{X}}$ in the sense of Blackwell (1951).

$\hat{f}_{\boldsymbol{Y}}$ in (11) has the following property desirable in any rankings of information structures:

$$\frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}} = E\left[\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\middle|\boldsymbol{Y} = \boldsymbol{y}\right], \tag{12}$$

where the expectation is over the joint density$\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a)$.[10]

From the equation in (12) and the proof of Corollary 2, the optimal contract $s_f(\boldsymbol{y})$ is characterised as follow:

- $s_f(\boldsymbol{y})$ is a fixed payment or a penalty contract,

where the joint density is expressed as $\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})f(\theta, \boldsymbol{x}|a)$. Thus, full disclosure (disclosure of the non-contractible signal $\boldsymbol{X}$) will improve the agent's action only if it increase implicit incentives. With regard to the impact on implicit incentives, one can refer the results of Dewatripont et al. (1999).

[10] Blackwell theorem is a sufficient condition for Kim's MPS criterion. The fact is proved by Kim (1995).

Lemma 2. (Dewatripont et al. (1999), Lemma 5.1) *If for a statistic $\boldsymbol{Y}$, $\hat{f}_a(\boldsymbol{x}|a)/\hat{f}(\boldsymbol{x}|a)$ and θ are conditionally positively correlated (respectively, negatively correlated) given $\boldsymbol{Y}$, implicit incentives are greater (respectively lower) when the market has information $\boldsymbol{X}$ than when the market has information $\boldsymbol{Y}$.*

Proof of Lemma 1 is found in Dewatripont et al. (1999).

Lemma 2 implies that the disclosure of $\boldsymbol{X}$ can either increase or decrease implicit incentives. Then, this section concerns the properties of the information structure of $(\boldsymbol{X}, \boldsymbol{Y}, \theta)$ that makes the full disclosure beneficial (or detrimental). This section proves the following proposition:

Proposition 3. *If $\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a)$ has the MLRP and if $(\boldsymbol{Y}, \boldsymbol{X}, \theta)$ is CIS, then the implicit incentive is greater when the market has information $\boldsymbol{X}$ than $\boldsymbol{Y}$. Similarly, if $\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|-a)$ have the MLRP and if $(\boldsymbol{Y}, \boldsymbol{X}, \theta)$ is CIS, then the implicit incentive is weaker when the market has information $\boldsymbol{X}$ than $\boldsymbol{Y}$.*

The proof is in Appendix.

Proposition 3 generalises Proposition 5.2 of Dewatripont et al. (1999) to signal vectors.

6 Conclusion

This paper adds contractible signals to the general-distribution career concern model of Dewatripont et al. (1999) and provides some properties of the information structure where the interplay between contracts and ceareer concerns are considered. This paper provides new insights into the literature on information system rankings in a principal-agent model.

Appendix

Proof of Proposition 1.

Proof. The proof is similar to Lemma 1 of Jewitt (1988). Substituting (5) into (3), one obtains

$$\iint U(w)\Big(\frac{\lambda}{U'(w)}-1\Big)\hat{f}d\boldsymbol{x}d\boldsymbol{y}=\mu V'(a). \tag{13}$$

Using the fact that $E\Big[\frac{\hat{f}_a}{\hat{f}}\Big]=0$, (5) gives

$$E\Big[\frac{\lambda}{U'(w)}\Big]=1. \tag{14}$$

(14) implies that the covariance of $U(w)$ and $\frac{\lambda}{U'(w)}$ is equal to $\mu V'(a)$. Since $U(w)$ and $\frac{\lambda}{U'(w)}$ are monotone increasing and from Assumption 3 $(\boldsymbol{X},\boldsymbol{Y})$ is implied to be association, their covariance is positive. Since $V'(a)$ is positive, one knows that $\mu>0$. □

Proof of Proposition 2.

Proof. $\hat{f}(\boldsymbol{x},\boldsymbol{y}|a)$ can be rewritten as

$$\begin{aligned}\hat{f}(\boldsymbol{x},\boldsymbol{y}|a)&=\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x},a)\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a),\\&=\hat{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y},a)\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}|a),\end{aligned}$$

where $\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x},a)$ is a density function of $\boldsymbol{y}$ conditioned by $(\boldsymbol{x},a)$, and $\hat{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y},a)$ is a density function of $\boldsymbol{x}$ conditioned by $(\boldsymbol{y},a)$. Since

$$\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_a}{\hat{f}}\right)=\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}}\right),$$

and from Observation 1 and Proposition 1 one knows that $\frac{d}{d\boldsymbol{y}}\left(\frac{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}}\right)$ and $s'_f(\boldsymbol{y})$ are of the same sign. $\frac{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}}$ can be written as

$$\frac{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}}=\frac{\hat{f}_{\boldsymbol{X}|\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{X}|\boldsymbol{Y}}}+\frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}-\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}.$$

Thus,

$$cov(\frac{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{Y}|\boldsymbol{X}}},y_i)=cov(\frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}},y_i)-cov(\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}},y_i),\ \ for\ all\ \ i=1,\ldots,n.$$

If there is $y_i, i = 1, \ldots, n$, such that $cov(\frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}, y_i) \leq cov(\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}, y_i)$ holds, the above equation tells that $\tilde{s}'_f(\boldsymbol{y}) > 0$ have to be denied. Hence, $s_f(\boldsymbol{y})$ violates non-decreasing monotonicity or $s_f(\boldsymbol{y})$ is fixed payment.

Conversely, if there is $y_i, i = 1, \ldots, n$, such that $\frac{\partial s_f(\boldsymbol{y})}{\partial y_i} \leq 0$ holds, then $cov(\frac{\hat{f}_{\boldsymbol{Y}a}}{\hat{f}_{\boldsymbol{Y}}}, y_i) \leq cov(\frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}, y_i)$ holds. This complete the proof of Proposition 2. $\square$

To prove Proposition 3, the following lemma must first be proved.

Lemma 3. *If $(\boldsymbol{Y}, \boldsymbol{X}, \theta)$ is CIS, then $(\boldsymbol{X}, \theta | \boldsymbol{Y} = \boldsymbol{y})$ is associated.*

Proof. Write $Cov_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}[\phi(\boldsymbol{X}, \theta), \psi(\boldsymbol{X}, \theta)] = E_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}\phi\psi - E_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}\phi \cdot E_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}\psi$, where the argument of ϕ and ψ are omitted and where $E_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}$ denotes expectation over the conditional distribution of $(\boldsymbol{X}, \theta)$ given $\boldsymbol{Y}$. Using the notation $Cov_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}[\phi, \psi]$ can be rewritten as

$$
\begin{aligned}
Cov_{(\boldsymbol{X},\theta)|\boldsymbol{Y}}[\phi, \psi] &= E_{\boldsymbol{X}|\boldsymbol{Y}} E_{\theta|\boldsymbol{X},\boldsymbol{Y}}\phi\psi - E_{\boldsymbol{X}|\boldsymbol{Y}} E_{\theta|\boldsymbol{X},\boldsymbol{Y}}\phi \cdot E_{\boldsymbol{X}|\boldsymbol{Y}} E_{\theta|\boldsymbol{X},\boldsymbol{Y}}\psi \\
&= E_{\boldsymbol{X}|\boldsymbol{Y}} E_{\theta|\boldsymbol{X}}\phi\psi - E_{\boldsymbol{X}|\boldsymbol{Y}} \left\{ E_{\theta|\boldsymbol{X}}\phi \cdot E_{\theta|\boldsymbol{X}}\psi \right\} \\
&\qquad + E_{\boldsymbol{X}|\boldsymbol{Y}} \left\{ E_{\theta|\boldsymbol{X}}\phi \cdot E_{\theta|\boldsymbol{X}}\psi \right\} - E_{\boldsymbol{X}|\boldsymbol{Y}} E_{\theta|\boldsymbol{X}}\phi \cdot E_{\boldsymbol{X}|\boldsymbol{Y}} E_{\theta|\boldsymbol{X}}\psi \\
&= E_{\boldsymbol{X}|\boldsymbol{Y}} Cov_{\theta|\boldsymbol{X}}[\phi, \psi] + Cov_{\boldsymbol{X}|\boldsymbol{Y}}[E_{\theta|\boldsymbol{X}}\phi, E_{\theta|\boldsymbol{X}}\psi].
\end{aligned}
$$

Now assume ϕ, ψ non-decreasing. Then, $Cov_{\theta|\boldsymbol{X}}[\phi, \psi] \geq 0$ by the fact that the set of consisting of a single random variable is associated, and so $E_{\boldsymbol{X}|\boldsymbol{Y}} Cov_{\theta|\boldsymbol{X}}[\phi, \psi] \geq 0$. To show $Cov_{\boldsymbol{X}|\boldsymbol{Y}}[E_{\theta|\boldsymbol{X}}\phi, E_{\theta|\boldsymbol{X}}\psi] \geq 0$, first show that $E_{\theta|\boldsymbol{X}=\boldsymbol{x}}\phi$ and $E_{\theta|\boldsymbol{X}=\boldsymbol{x}}\psi$ are non-decreasing in $\boldsymbol{x}$. From assumption, $(\boldsymbol{Y}, \boldsymbol{X}, \theta)$ is CIS, which implies $(\boldsymbol{X}, \theta)$ is CIS by the fact that $\boldsymbol{X} = \boldsymbol{x}$ reveals the realisation of $\boldsymbol{Y}$, which follows that $E_{\theta|\boldsymbol{X}=\boldsymbol{x}}\phi$ and $E_{\theta|\boldsymbol{X}=\boldsymbol{x}}\psi$ are non-decreasing in $\boldsymbol{x}$. Second, show that $(\boldsymbol{X}|\boldsymbol{Y})$ is associated. From assumption, $(\boldsymbol{Y}, \boldsymbol{X}, \theta)$ is CIS, which implies $(\boldsymbol{Y}, \boldsymbol{X})$ is CIS, which implies

$$
\begin{aligned}
&P\left(X_i > x_i | \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X}_J = \boldsymbol{x}_J\right) \leq P\left(X_i > x_i | \boldsymbol{Y} = \boldsymbol{y}', \boldsymbol{X}_J = \boldsymbol{x}'_J\right) \\
&\qquad\qquad\qquad\qquad\qquad \textit{for all } \boldsymbol{x}_J \leq \boldsymbol{x}'_J,\ \boldsymbol{y} \leq \boldsymbol{y}',\ i = 2, \ldots, n,
\end{aligned}
$$

where $J = \{1, \ldots, i-1\}$, which finally implies $(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y})$ is CIS, and it follows that $Cov_{\boldsymbol{X}|\boldsymbol{Y}}[E_{\theta|\boldsymbol{X}}\phi, E_{\theta|\boldsymbol{X}}\psi] \geq 0$. □

Proof of Proposition 3.

Proof. By Lemma 2, the implicit incentive is greater (weaker) when the market has information $\boldsymbol{X}$ than $\boldsymbol{Y}$ if

$$Cov_{\boldsymbol{X},\theta|\boldsymbol{Y}}\left[\theta, \frac{\hat{f}_{\boldsymbol{X}a}}{\hat{f}_{\boldsymbol{X}}}\right] \geq (\leq) 0. \tag{15}$$

But by Lemma 3 $(\boldsymbol{X}, \theta|\boldsymbol{Y} = \boldsymbol{y})$ is associated, which implies that (15) is positive if $\frac{\hat{f}_{\boldsymbol{X}^a}(\boldsymbol{x}|a)}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|a)}$ has the MLRP and negative if $\frac{\hat{f}_{\boldsymbol{X}^a}(\boldsymbol{x}|-a)}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}|-a)}$ has the MLRP. □

References

Blackwell, David (1951) "Comparison of experiments," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, Vol. 1, pp. 93–102.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole (1999) "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, Vol. 66, No. 1, Special Issue: Contracts, pp. 183–198.

Fama, Eugene F. (1980) "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, Vol. 88, No. 2, pp. 288–307.

Gibbons, Robert and Kevin J. Murphy (1992) "Optimal Incentive contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, Vol. 100, No. 3, pp. 468–505.

Gjesdal, Frøystein (1982) "Information and incentives: The agency information problem," *Review of Economic Studies*, Vol. 49, No. 3, pp. 373–390.

Grossman, Sanford J. and Oliver D. Hart (1983) "An Analysis of the Principal-Agent Problem," *Econometrica*, Vol. 51, No. 1, pp. 7–45.

Holmström, Bengt (1979) "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. 10, No. 1, pp. 74–91.

——— (1999) "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, Vol. 66, No. 1, pp. 169–182.

Jewitt, Ian (1988) "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, Vol. 56, No. 5, pp. 1177–1190.

Kaarbøe, Oddvar M. and Trond E. Olsen (2008) "Distorted Performance Measures and Dynamic Incentives," *Journal of Economics & Management Strategy*, Vol. 17, No. 1, pp. 149–183.

Kim, Son Ku (1995) "Efficiency of an Information System in an Agency Model," *Econometrica*, Vol. 63, No. 1, pp. 89–102.

Meyer, Margaret A. and John Vickers (1997) "Performance comparisons and dynamic incentives," *Journal of Political Economy*, Vol. 105, No. 3, pp. 547–581.

Müller, Alfred and Marco Scarsini (2001) "Stochastic Comparison of Random Vectors with a Common Copula," *Mathematics of operations research*, Vol. 26, No. 4, pp. 723–740.

Müller, Alfred and Dietrich Stoyan (2002) *Comparison Methods for Stochastic Models and Risks*, Chichester: John Wiley & Sons.

Ogaku, Michiko (2014) "Managerial incentive problems: A role of multi-signals." Discussion paper, No. 1319, Department of Social Systems and Management, University of Tsukuba.

Rogerson, William P. (1985) "The First-Order Approach to Principal-Agent Problems," *Econometrica*, Vol. 53, No. 6, pp. 1357–1367.

Sabac, Florin (2007) "Dynamic agency with renegotiation and managerial tenure," *Management science*, Vol. 53, No. 5, pp. 849–864.

——— (2008) “Dynamic incentives and retirement,” *Journal of Accounting and Economics*, Vol. 46, No. 1, pp. 172–200.

Shaked, Moshe and J. George Shanthikumar (2007) *Stochastic Orders*: Springer.

Sinclair-Desgagné, Bernard (1994) “The First-Order Approach to Multi-Signal Principal-Agent Problems,” *Econometrica*, Vol. 62, No. 2, pp. 459–465.