

**Efficiency of an Information System in an Agency Model**

Son Ku Kim

University of California, Los Angeles

UCLA Department of Economics  
Working Paper #608S  
September 1990

This paper is drawn from my UCLA Doctoral Dissertation. The issue was introduced to me by Yoon Suh. I am especially indebted to my advisor, John Riley, for his continuous advice and encouragement. Also the comments and suggestions of David Levine, Seonghwan Oh, and particularly Yoon Suh are gratefully acknowledged.

## **Abstract**

Different information systems are compared in terms of their relative efficiencies in an agency model. It is shown that the variance of the likelihood ratio distribution ranks such information systems when the agent's utility function is a square root. Also the second order stochastic dominance relation between the likelihood ratio distributions derived from the original information systems is found to be sufficient to rank information systems under quite general assumptions about the agent's utility function. Furthermore, it is shown that the second order stochastic dominance criterion is equivalent to Holmstrom's informativeness criterion when the comparison is done for inclusive information systems.

## 1 Introduction

A principal-agent problem arises when it is impossible for a principal to design a reward scheme which is conditional upon an agent's choice of an action because of the principal's lack of ability to observe that action. In the basic agency model, there is a set of common observables imperfectly correlated with the agent's hidden action choice and the principal's interest in controlling the agent's decision-making leads him to utilize those observables. The system that generates those observables is called an information system.

Most often it is assumed that the principal observes the output produced by the agent, either exactly or with error.<sup>1</sup> More generally, in a multi-agent setting, an agent's contribution to aggregate output may be very difficult to estimate. In such cases, the principal has an incentive to monitor some facets of an agent's input. The information service is then one or more signals from a monitoring technology, where each signal is correlated with the agent's input.

Even though growing attention has been paid to the modelling of the principal-agent relationship over the last decade, this literature is primarily focused on the problem of designing an incentive contract for some given information system.<sup>2</sup> However, in reality, the principal typically must also choose among alternative information systems before designing the incentive scheme. Thus the principal's problem of designing an incentive contract through which he wishes to control the agent's decision-making consists of two stages: the choice of an information system and the design of an incentive scheme based on the information system that is chosen in the first stage.

---

<sup>1</sup>In this traditional case where the only observable is the output, the production technology can play a role of information system that gives a signal correlated with the agent's hidden action.

<sup>2</sup>See Ross (1973), Mirrlees (1974), Harris and Raviv (1979), Holmstrom (1979,1982), Shavell (1979), Grossman and Hart (1983), among others.

This paper focuses on the first stage. To illustrate the issues, consider the following simple bucolic example. A risk neutral ranch holder (principal) hires a risk averse cowhand (agent) to manage a herd of milch cows. The principal must choose between two breeds of cows which are equally costly. The amount of milk produced in a given period is affected not only by the cowhand's effort that is hidden to the ranch holder but also by the weather which changes stochastically. It is assumed that the expected amounts of milk from both breeds are the same if combined with the cowhand's same effort but the cow of breed *A* is more weather resistant than that of breed *B*. Thus the amount of milk from cow *A* is less sensitive to the change of the weather. Consequently, conditional on any particular effort level of the cowhand, the distribution of the amount of milk from cow *A* is less risky.<sup>3</sup> The first problem to be solved by the principal is which breed should be given to the cowhand.<sup>4</sup> A naive answer is that the ranch holder will be indifferent between the two breeds since being risk neutral he is only concerned with the expected amount of milk. Of course this is incorrect because not only the expected amount of milk but also the cost needed to control the cowhand's hidden action is important in this moral hazard situation. For any effort level of the cowhand, if the variance of milk output is greater for breed *B* then the information contained in the output signal is less revealing to the ranch holder. This suggests that the ranch holder will prefer the less risky breed *A* since, for a given payment scheme based on milk output, he imposes less risk on the cowhand.

While this intuition is sound, as far as it goes, it is incomplete except in special cases, for example, where milk output is normally distributed.<sup>5</sup> In this paper we establish that the key factor in ranking information systems in an agency model is not the distribution of the signal conditional upon the agent's action but the derived

---

<sup>3</sup>The term 'less risky distribution' here means a distribution with a less variance.

<sup>4</sup>This bucolic example can easily be extended to the real economy where the principal provides a specific production project among many alternatives and the agent's provides hidden effort.

<sup>5</sup>For this issue, see Kim and Suh (1990).

*likelihood hood ratio distribution* of the signal.

The related subject of ranking information systems of *a decision maker* was initiated by Blackwell (1951, 1953) and the economic comparability of such ranking was developed by Marschak and Miyazawa (1968). More recently there have been two efforts to analyze the ranking of information systems in an agency model. The first result regarding the choice of an information system is the informativeness criterion in Holmstrom (1979). This criterion states that, between any two costless public information systems having an inclusive relation,<sup>6</sup> the information system with additional signals is strictly preferred by the principal if and only if at least one of those additional signals is informative about the agent's action choice. Holmstrom's notion of informativeness is defined in terms of a sufficient statistic.<sup>7</sup> Second, Grossman and Hart (1983) and Gjesdal (1982) apply Blackwell's ranking information systems to the agency model and they show that the sufficient part of Blackwell's theorem works in the agency model, that is, when two information systems satisfy Blackwell's sufficiency condition they can be ranked in their efficiencies in the agency model.

However, the informativeness concept is very restrictive as a ranking criterion in the sense that it is impossible to apply such a criterion for the ranking of information systems which cannot be ordered in terms of a sufficient statistic as the bucolic example mentioned above. Moreover, the application of Blackwell's theorem to the agency model still remains obscure in many respects. The purpose of this paper is to find another criterion for the ranking of information systems in the agency model which can be applied to a broader set of information systems and, in doing so, shed light on the earlier work.

The paper is organized as follows: In Section 2 the basic agency model is formu-

---

<sup>6</sup>An inclusive relation between two information systems here means that one information system has additional sources of signals than the other. However, the inclusive relation generally means that one information system is a finer partition of the other.

<sup>7</sup>Holmstrom (1982) extends his result to the generalized notion of a sufficient statistic.

lated and in Section 3 two criteria for the ranking of information systems are derived and discussed. First, by imposing a restriction on the agent's preference we obtain a measure by which we can rank all information systems. That is, when the agent's utility function is the square root function, the variance of the likelihood ratio becomes the measure which ranks all the information systems if such variances exist (**Variance Criterion**). Second, under the general additive separability assumption for the agent's utility function, we show that one information system is more efficient than the other if its derived likelihood ratio distribution is second order stochastically dominated by that of the other information system (**SOSD Criterion: Second Order Stochastic Dominance Criterion**). So it is not the distribution of an information system itself but the distribution of its derived likelihood ratio that is important in deciding the relative efficiency of the information system.<sup>8</sup>

In Section 4, it is shown that the SOSD criterion includes Holmstrom's informativeness criterion in the sense that it becomes equivalent to the informativeness criterion when applied to the ranking of inclusive information systems. In Section 5, it is shown that the sufficient part of Blackwell's ranking of information systems can be applied to the agency model and the basic difference between Blackwell's theorem and the SOSD criterion is explored. In Section 6, the welfare implication of moral-hazard is investigated.

In Section 7, as concluding remarks, we briefly mention further extensions and possible applications.

---

<sup>8</sup>The reason the distribution of the likelihood ratio is more important than the distribution itself is that the former can clearly represent the role of the agent's hidden action choice in the information system which is important in deciding the efficiency of the information system.

## 2 Basic Model

We consider a two-person, single-period agency model. The agent participates in this decentralized economy by providing some productive effort  $a \in [\underline{a}, \bar{a}] \subset R$ , not observed by the principal. There is a signal vector  $y$  commonly observed by the principal and the agent without cost. The signal vector  $y = Y(a, \theta) \in Y \subset R^m, m \geq 1$ , generated from the information system  $Y$ , is received after the agent has made his choice of  $a$  and the nature its choice of  $\theta \in \Theta$ . The monetary outcome  $x$  is determined by the stochastic production process  $X$ , i.e.,  $x = X(a, \theta) \in X$ . The monetary outcome  $x$  may be included in  $y$  or may not.<sup>9</sup> The agent's compensation  $s$  is a function of the jointly observed signal  $y$ .

Given a probability measure on  $\Theta$ , the information technology  $Y$  can be described by the joint density function  $h(y|a)$  parameterized by the agent's effort. Thus,  $h(y|a) \in \Gamma$  can be interpreted as an information system which generates a signal vector  $y$  informative about  $a$  where  $\Gamma$  denotes a set of information systems to be ranked.<sup>10</sup>

### Assumption 1:

The principal is assumed to be risk neutral while the agent is work- and income risk-averse. Furthermore, the agent's preference on wealth and effort is assumed to be additively separable, i.e.,

$$U(s, a) = u(s) - v(a) \quad u' > 0, u'' < 0, v' > 0$$

---

<sup>9</sup>In much of the moral hazard literature, the monetary outcome  $x$  is assumed to be the only common observable. In this case, the signal  $y = x$ . In general, however, the outcome  $x$  will not be observed without cost. On the other hand, other signals than the monetary outcome may be more easily observed. See Gjesdal (1982).

<sup>10</sup>This parameterized distribution formulation is pioneered by Mirrlees (1974) and further explored by Holmstrom (1979).

where  $v$  denotes the agent's disutility of effort.

**Assumption 2:**

$$\frac{\partial \int^y h(t|a)dt}{\partial a} < 0, \quad \forall h \in \Gamma, \forall y \in Y.$$

**Assumption 3:**

For any value  $y$ ,  $\forall h(y|a) \in \Gamma$  is twice differentiable in  $a$ .

**Assumption 4:**

$Y$ , the image of information system  $Y$ , is independent of the agent's action choice.

**Assumption 2** implies that the signal  $y$  is increasing in  $a$  in the sense of first order stochastic dominance while **Assumption 3** is provided to characterize the optimal incentive scheme. However, **Assumption 3** also rules out perfect information systems from  $\Gamma$ . Thus, the focus of this paper is to rank imperfect information systems. **Assumption 4** also eliminates perfect information systems. But more strongly it eliminates all the moving support cases from  $\Gamma$ .<sup>11</sup>

The agency problem with an information system  $h(y|a) \in \Gamma$  is the following:

$$\begin{aligned} \text{Max} \quad & \int [x - s(y)]h(y|a)dy \\ & s(y) \geq k, a \\ \text{s.t} \quad & \text{(i) } \int u(s(y))h(y|a)dy - v(a) \geq \bar{U} \\ & \text{(ii) } a \in \text{argmax} \int u(s(y))h(y|a)dy - v(a) \end{aligned}$$

where  $\bar{U}$  denotes the agent's reservation level of utility. Here the lower bound of the agent's compensation,  $k$ , is provided to avoid a nonexistence problem. In practice,  $k$

<sup>11</sup>The reason we eliminate all the moving support cases from  $\Gamma$  is a technical one to guarantee  $\int \frac{h}{k}(y|a)h(y|a)dy = 0, \forall h \in \Gamma$ .



can be interpreted as the subsistence level of the agent's income. The first and the second constraints are referred to as the participation constraint and the incentive compatibility constraint, respectively.

Of course, the comparison of information systems in  $\Gamma$  in terms of their values will be accomplished by solving the above program for every information system in  $\Gamma$  and then by ranking in terms of the principal's expected payoff.<sup>12</sup> However, this ranking will be sensitive to the characteristics of the agency problem especially to the agent's utility function. Our purpose, however, is to find a ranking criterion that is insensitive to the characteristics of the agency problem. Furthermore, our main interest lies in deciding the efficiency of an information system which is associated with the precision of the information system. Thus, it is natural to make the following assumption about  $\Gamma$ .

**Assumption 5:**

Every information system  $h \in \Gamma$  has the same expected monetary outcome for any given  $a$ , i.e.,

$$\int xh(y|a)dy = E(X|a), \quad \forall h \in \Gamma, \quad \forall a \in [a, \bar{a}]$$

**Assumption 5** means every information system in  $\Gamma$  has the same expected productivity, thus we ignore the productivity aspect on the efficiency of an information system. So ranking those information systems can be equivalently done by comparing the principal's minimized expected compensation costs under such information systems.

On the other hand, the changes in agency's welfare with respect to changes in information system will be generated through two different but mixed ways, i.e., the

---

<sup>12</sup>The concept of value here is that of the principal's economic demand price.

changes in optimal incentive scheme and the changes in optimal action choice. However, our focus here is further restricted to ranking information systems in inducing the same action. Thus, the agent's action choice is fixed and our concern is that which information system is better (less costly) in inducing the same action.

In order to induce a given action  $a$  at a minimum cost the principal has to solve the following program under information system  $h(y|a)$ :<sup>13</sup>

$$\begin{aligned} & \text{Min} \quad \int s(y)h(y|a)dy \\ & s(y) \geq k \\ & \text{s.t} \quad \text{(i) } \int u(s(y))h(y|a)dy - v(a) \geq \bar{U} \\ & \quad \quad \text{(ii) } a \in \text{argmax} \int u(s(y))h(y|a)dy - v(a) \end{aligned}$$

Now by assuming Holmstrom's (1979) first order approach is valid,<sup>14</sup> the second constraint can be reduced to

$$\text{(ii')} \quad \int u(s(y))h_a(y|a)dy - v'(a) = 0$$

where  $h_a(y|a)$  and  $v'(a)$  denote the first derivatives with respect to  $a$ .

Then the cost-minimizing contract  $s_h^a(y)$  in inducing the agent a given action choice  $a \neq \underline{a}$  under the information system  $h(y|a)$  must satisfy

$$\frac{1}{u'(s_h^a(y))} = \lambda_h + \mu_h \frac{h_a}{h}(y|a) \tag{1}$$

---

<sup>13</sup>Thus, we ignore the welfare implication due to the possible changes in optimal action choice followed by the changes in information system.

<sup>14</sup>This assumption restricts  $\Gamma$ . Rogerson (1985) shows that MLRP (Monotone Likelihood Ratio Property) and CDFC (Convexity of the Distribution Function Condition) are sufficient for the validity of the first order approach. As noted many times (See Hart and Holmstrom (1985), Jewitt (1988).), most of the distributions commonly used in statistics and economics do not satisfy CDFC. However, Jewitt (1988) recently finds a less restrictive condition for the validity of the first-order approach and shows that many well-known families of distribution functions, e.g., exponential, gamma, chi-square, Poisson, satisfy his condition.

for almost every  $y$  for which (1) has a solution  $s_h^a(y) \geq k$  and otherwise  $s_h^a(y) = k$ .  $\lambda_h$  and  $\mu_h$  denote the Lagrangian multipliers of constraints (i) and (ii') respectively, given information system  $h(y|a)$ .

**Observation 1:**

The Lagrangian multipliers of constraints (i) and (ii'),  $\lambda_h$  and  $\mu_h$ , are both strictly positive.

It is well known that when both the principal and the agent possess homogeneous beliefs at the time of contracting, the agent always gets his reservation utility at the optimum (i.e.,  $\lambda_h > 0$ ). Also, as shown by Holmstrom (1979), we cannot attain the first-best solution (i.e.,  $\mu_h > 0$ ).

**3 Analysis**

Since, from **Observation 1**, the agent always receives his reservation utility at the optimum under any information system  $h \in \Gamma$ , the relative efficiency of the information system in inducing a given action is determined by its relative expected compensation cost in enforcing the agent such an action choice. Consider two information systems  $f$  and  $g$ ,  $f, g \in \Gamma$ .

**Definition 1:** (Efficiency for a given action)

Information system  $f(y|a)$  is 'more efficient at  $a$ ' than  $g(y|a)$  if and only if, at the optimum, the expected compensation for inducing the action choice  $a$  under  $f(y|a)$  is less than that under  $g(y|a)$ , that is,

$$Es_f^a < Es_g^a$$

where,

$$Es_f^a \equiv \int s_f^a(y)f(y|a)dy$$

$$Es_g^a \equiv \int s_g^a(y)g(y|a)dy$$

**Proposition 1:**

No information system  $h \in \Gamma$  is ‘more efficient at  $\underline{a}$ ’ than others in  $\Gamma$ .

**Proof:**

Since  $\underline{a}$  is the least costly action to the agent (i.e.,  $v' > 0$ ), to induce the agent to choose that action for any information system it is enough to give the agent a constant wage which guarantees the agent  $\bar{U}$ . Thus the principal becomes indifferent about the information systems. ■

**Definition 2:** (Efficiency)

Information system  $f(y|a)$  is ‘more efficient’ than  $g(y|a)$  if and only if  $Es_f^a \leq Es_g^a$  for all  $a \in (\underline{a}, \bar{a}]$ , the strict inequality holding for some  $a$ .

**3.1 Variance Criterion**

We start with a simple example by specifying the agent’s utility function a square root because this simple example provides many clear intuitions as a benchmark. Since the purpose of this presentation is just to show such intuitions, we temporarily assume without much loss of generality that (1) is satisfied for almost every  $y$  under any information system  $h \in \Gamma$ .

**Proposition 2 (Equivalence Theorem):**<sup>15</sup>

When the agent’s utility function is a square root, the following equivalences hold for any given  $a \in (\underline{a}, \bar{a}]$ :

---

<sup>15</sup>The same result is used in Kim and Suh (1990).

$$\begin{aligned}
\text{var}\left(\frac{f_a}{f}\right) > \text{var}\left(\frac{g_a}{g}\right) &\iff \mu_f < \mu_g \\
&\iff \text{var}\left[\frac{1}{u'(s_f^a)}\right] < \text{var}\left[\frac{1}{u'(s_g^a)}\right] \\
&\iff Es_f^a < Es_g^a
\end{aligned}$$

**Proof:**

When the agent has a square root utility function  $u(s) = 2\sqrt{s}$ , by (1) we get

$$Es_h^a = \int (\lambda_h + \mu_h \frac{h_a}{h})^2 h(y|a) dy, \quad h = f, g \quad (2)$$

We now use the fact that two constraints (i) and (ii') are binding.

From (i),

$$\int (\lambda_h + \mu_h \frac{h_a}{h}) h(y|a) dy = \frac{\bar{U} + v(a)}{2} \equiv \frac{C}{2}, \quad h = f, g \quad (3)$$

and, from (ii')

$$\int (\lambda_h + \mu_h \frac{h_a}{h}) \frac{h_a}{h} h(y|a) dy = \frac{v'(a)}{2} \equiv \frac{D}{2}, \quad h = f, g \quad (4)$$

By (3), since  $\int \frac{f_a}{f} f(y|a) dy = \int \frac{g_a}{g} g(y|a) dy = 0$ ,

$$\lambda_f = \lambda_g = \frac{C}{2}.$$

And by (4),

$$\mu_f \int \left(\frac{f_a}{f}\right)^2 f(y|a) dy = \mu_g \int \left(\frac{g_a}{g}\right)^2 g(y|a) dy = \frac{D}{2}. \quad (5)$$

Since  $D$  is a constant that is exogenously given,  $\text{var}\left(\frac{f_a}{f}\right) [\equiv \int \left(\frac{f_a}{f}\right)^2 f(y|a) dy] > \text{var}\left(\frac{g_a}{g}\right)$  is equivalent to  $\mu_f < \mu_g$ . From (1),

$$E\left[\frac{1}{u'(s_h^a)}\right] \equiv \int (\lambda_h + \mu_h \frac{h_a}{h}) h(y|a) dy = \lambda_h.$$

So, by (1) and (5),

$$\text{var}\left[\frac{1}{u'(s_f^a)}\right] = \mu_f^2 \int \left(\frac{f_a}{f}\right)^2 f(y|a) dy = \mu_f^2 \text{var}\left(\frac{f_a}{f}\right) = \mu_f \frac{D}{2}$$

$$\text{var}\left[\frac{1}{u'(s_g^a)}\right] = \mu_g^2 \int \left(\frac{g_a}{g}\right)^2 g(y|a) dy = \mu_g^2 \text{var}\left(\frac{g_a}{g}\right) = \mu_g \frac{D}{2}$$

Therefore, since  $D$  is positive,  $\mu_f < \mu_g$  is equivalent to  $\text{var}\left[\frac{1}{u'(s_f^a)}\right] < \text{var}\left[\frac{1}{u'(s_g^a)}\right]$ .

Furthermore, from (2),

$$Es_f^a = \lambda_f^2 + \mu_f^2 \text{var}\left(\frac{f_a}{f}\right) = \frac{C^2}{4} + \text{var}\left[\frac{1}{u'(s_f^a)}\right]$$

$$Es_g^a = \lambda_g^2 + \mu_g^2 \text{var}\left(\frac{g_a}{g}\right) = \frac{C^2}{4} + \text{var}\left[\frac{1}{u'(s_g^a)}\right]$$

So,  $\text{var}\left[\frac{1}{u'(s_f^a)}\right] < \text{var}\left[\frac{1}{u'(s_g^a)}\right]$  is equivalent to  $Es_f^a < Es_g^a$ . Therefore, the information system  $f(y|a)$  is less costly than  $g(y|a)$  if and only if  $\text{var}\left(\frac{f_a}{f}\right) > \text{var}\left(\frac{g_a}{g}\right)$ . ■

Statistically,  $\frac{h_a}{h}(y|a)$  is the derivative of the log likelihood function  $\log h(y|a)$  with respect to the unknown parameter  $a$ . More precisely, in the two action case ( $a_L, a_H$ ), where  $a_L$  and  $a_H$  denote the agent's low effort and high effort respectively, the term  $\frac{h_a}{h}(y|a)$  reduces to  $1 - \frac{h(y|a_L)}{h(y|a_H)}$ . Since  $\frac{h(y|a_L)}{h(y|a_H)}$  represents the likelihood ratio of the agent's low effort choice when  $y$  is observed,  $1 - \frac{h(y|a_L)}{h(y|a_H)}$  can be thought as the likelihood ratio of the agent's high effort choice when  $y$  is observed. Therefore,  $\frac{h_a}{h}(y|a)$ , which is the counterpart of  $1 - \frac{h(y|a_L)}{h(y|a_H)}$  in the continuous action case, also represents the likelihood ratio of the the agent's action choice  $a$ . In other words, the term  $\frac{h_a}{h}(y|a)$  designates the principal's statistical inference about the agent's action choice  $a$  when  $y$  is observed. Since the first best risk sharing rule between the risk neutral principal and the risk averse agent is constant regardless of  $y$ , the fact that the optimal incentive scheme  $s_h^a(y)$  should be designed in the way that is shown in

equation (1) in the presence of moral-hazard implies that the term  $\frac{h_a}{h}(y|a)$ , which together with  $\mu_h$  denotes the deviation from the first best risk sharing rule, is needed for incentive provision in designing an incentive scheme.<sup>16</sup>

**Proposition 2** indicates that when the agent has a square root utility function, all information systems in  $\Gamma$  can be ranked by the magnitude of  $var(\frac{h_a}{h})$ . That is,  $var(\frac{h_a}{h})$  provides a total ordering of the information systems for such an action choice when the agent's utility is a square root. The intuition that  $var(\frac{h_a}{h})$  is related to the relative efficiency of an information system  $h(y|a)$  is the following: In statistics,  $var(\frac{h_a}{h})$  is referred to as 'the amount of information' on 'a' contained in  $y$  under information system  $h(y|a)$ .<sup>17</sup> So, as shown in the equivalence theorem, when the principal can obtain more information about the agent's hidden action choice under information system  $f(y|a)$  (i.e.,  $var(\frac{f_a}{f}) > var(\frac{g_a}{g})$ ), the extent of the moral hazard problem is reduced under  $f(y|a)$  (i.e.,  $\mu_f < \mu_g$ ). And this reduction in the extent of the moral hazard problem is accompanied by the smaller deviation from the first best risk sharing rule which is needed for incentive provision, i.e.,  $var[\frac{1}{u'(s_f^*)}] < var[\frac{1}{u'(s_g^*)}]$ . Consequently, the principal can reduce his expected compensation cost by being able to give an incentive scheme closer to the first best.

**Corollary 1:**

When the agent's utility function is a square root, information system  $f(y|a)$  is 'more efficient' than  $g(y|a)$  if and only if  $var(\frac{f_a}{f}) \geq var(\frac{g_a}{g})$  holds for all  $a \in (\underline{a}, \bar{a}]$ , the strict inequality holding for some  $a$ .

---

<sup>16</sup>The first best risk sharing rule here means the one that obtains when the principal observes the agent's action choice directly. In this case, the optimal risk sharing rule will be constant.

<sup>17</sup>See Freeman (1963).

### 3.2 Second Order Stochastic Dominance (SOSD) Criterion

Our main purpose, however, is not to find a ranking criterion which is appropriate for one specific utility function of the agent but to find the ranking criterion which can be applied to the agent's general utility function. So a natural question whether the variance criterion can still be applied to the agent's general utility function case will be answered here.

#### Definition 3: ( Weak SOSD )

The distribution function  $F$  'weakly dominates' over  $G$  in the sense of second order stochastic dominance if and only if

$$\int^y F(t)dt \leq \int^y G(t)dt \quad \text{for all } y \in Y.$$

#### Definition 4: ( (Strong) SOSD )<sup>18</sup>

The distribution function  $F$  '(strongly) dominates' over  $G$  in the sense of second order stochastic dominance if and only if

$$\int^y F(t)dt \leq \int^y G(t)dt \quad \text{for all } y \in Y.$$

and the strict inequality holds for at least one  $y \in Y$ .

#### Definition 5:(Likelihood Ratio Distribution Function)

$H_h^a(z)$  denotes the likelihood ratio distribution function of an action choice 'a' under information system  $h(y|a)$  where  $z = \frac{h_a}{h}(y|a)$ . That is,

$$H_h^a(z) = \Pr_h[Z \leq z] = \Pr\left[\frac{h_a}{h}(Y|a) \leq z\right] = \int_{\frac{h_a}{h}(y|a) \leq z} h(y|a)dy.$$

#### Proposition 3:

Information system  $f(y|a)$  is 'more efficient at  $a \neq \underline{a}$ ' than  $g(y|a)$  with any risk

<sup>18</sup>For details of SOSD, see Hadar and Russel (1969, 1971) and Hanoch and Levy (1969).



averse agent satisfying **Assumption 1** if the likelihood ratio distribution of that action choice under information system  $f(y|a)$ ,  $H_f^a(z)$ , is second order stochastically dominated by that under information system  $g(y|a)$ ,  $H_g^a(z)$ .

**Proof:**

Let us define, under information system  $h(y|a)$ ,  $h = f, g$

$$q_h \equiv q_h(y|a) = \lambda_h + \mu_h \frac{h_a}{h}(y|a)$$

Then,  $s_h^a(y)$  can be written as  $s(q_h)$  since the functional form of  $s$  with respect to  $q_h$  only depends upon the functional form of  $u$  with respect to  $s$  by (1). Also by (1)

$$q_h u'(s(q_h)) = \begin{cases} 1 & \text{when } q_h \geq \frac{1}{u'(k)} \\ q_h u'(k) & \text{when } q_h < \frac{1}{u'(k)} \end{cases} \quad (6)$$

since  $s(q_h) = k$  and  $u(s(q_h)) = u(k)$  when  $q_h < \frac{1}{u'(k)}$ .

Since the two constraints are always binding, i.e.,

$$(i) \int u(s(q_h)) h(y|a) dy = \bar{U} + v(a) \equiv C$$

$$(ii') \int u(s(q_h)) \frac{h_a}{h} h(y|a) dy = v'(a) \equiv D,$$

it follows that for any information system  $h(y|a)$ ,  $h = f, g$ ,

$$\begin{aligned} E(q_h u_h) &\equiv \int u(s(q_h)) q_h h(y|a) dy = \int u(s(q_h)) (\lambda_h + \mu_h \frac{h_a}{h}) h(y|a) dy \\ &= \lambda_h C + \mu_h D \end{aligned} \quad (7)$$

By dropping the subscript  $h$  for notational simplicity, now we know that  $s_q(q) \equiv \frac{\partial s(q)}{\partial q} \geq 0$ , since by (1) as  $q$  increases  $u'(s(q))$  is nonincreasing and by the concavity of the agent's utility function  $s(q)$  is nondecreasing. Let us define

$$\psi(q) \equiv s(q) - u(s(q))q \quad (8)$$

Then, using (6), we obtain when  $q \geq \frac{1}{u'(k)}$ ,

$$\frac{\partial\psi(q)}{\partial q} = s_q - u'(s(q))s_q q - u(s(q)) = -u(s(q))$$

$$\frac{\partial^2\psi(q)}{\partial q^2} = -u'(s(q))s_q < 0$$

and when  $q < \frac{1}{u'(k)}$ , since  $s_q = 0$ ,

$$\frac{\partial\psi(q)}{\partial q} = -u(k)$$

Thus,  $\psi(q)$  is continuous and concave in  $q$ .

Now define  $\bar{q} \equiv \bar{q}(y|a) = \lambda_f + \mu_f \frac{q_a}{g}(y|a)$ .

Using (8), we obtain

$$\begin{aligned} E\psi(\bar{q}) - E\psi(q_g) &\equiv \int [s(\bar{q}) - u(s(\bar{q}))\bar{q}]g(y|a)dy - \int [s(q_g) - u(s(q_g))q_g]g(y|a)dy \\ &= \int [s(\lambda_f + \mu_f \frac{q_a}{g}) - u(s(\lambda_f + \mu_f \frac{q_a}{g}))(\lambda_f + \mu_f \frac{q_a}{g})]g(y|a)dy \\ &\quad - \int [s(\lambda_g + \mu_g \frac{q_a}{g}) - u(s(\lambda_g + \mu_g \frac{q_a}{g}))(\lambda_g + \mu_g \frac{q_a}{g})]g(y|a)dy \end{aligned}$$

Since  $\psi(q)$  is concave in  $q$ , it is also concave in  $(\lambda, \mu)$  between  $(\lambda_f, \mu_f)$  and  $(\lambda_g, \mu_g)$ .

Therefore,

$$\begin{aligned} E\psi(\bar{q}) - E\psi(q_g) &\leq \{ \int [s_q(q_g) - u'(s(q_g))s_q(q_g)q_g]g(y|a)dy \} (\lambda_f - \lambda_g) \\ &\quad + \{ \int [s_q(q_g) - u'(s(q_g))s_q(q_g)q_g] \frac{q_a}{g} g(y|a)dy \} (\mu_f - \mu_g) \\ &\quad - \{ \int u(s(q_g))g(y|a)dy \} (\lambda_f - \lambda_g) + \{ \int u(s(q_g)) \frac{q_a}{g} g(y|a)dy \} (\mu_f - \mu_g) \} \end{aligned}$$

The first two terms in the right hand side become zero by (6) and the third term, by (7), reduces to  $-C(\lambda_f - \lambda_g) - D(\mu_f - \mu_g)$ . Thus by (7),

$$\begin{aligned}
E\psi(\bar{q}) - E\psi(q_g) &\leq -C(\lambda_f - \lambda_g) - D(\mu_f - \mu_g) \\
&= E(q_g u_g) - E(q_f u_f)
\end{aligned}$$

Using (8)

$$\begin{aligned}
Es_g^a - Es_f^a &= E\psi(q_g) - E\psi(q_f) + E(q_g u_g) - E(q_f u_f) \\
&\geq E\psi(\bar{q}) - E\psi(q_f) \\
&= \int [s(\lambda_f + \mu_f \frac{q_a}{g}) - u(s(\lambda_f + \mu_f \frac{q_a}{g}))](\lambda_f + \mu_f \frac{q_a}{g}) g(y|a) dy \\
&\quad - \int [s(\lambda_f + \mu_f \frac{f_a}{f}) - u(s(\lambda_f + \mu_f \frac{f_a}{f}))](\lambda_f + \mu_f \frac{f_a}{f}) f(y|a) dy
\end{aligned}$$

Thus, using the transformation in **Definition 5**, we can rewrite

$$Es_g^a - Es_f^a \geq \int w(z) dH_g^a(z) - \int w(z) dH_f^a(z) \quad (9)$$

where,

$$w(z) \equiv s(\lambda_f + \mu_f z) - u(s(\lambda_f + \mu_f z))(\lambda_f + \mu_f z) \quad (10)$$

Since  $\mu_h > 0$ ,  $w(z) \equiv \psi(\lambda_f + \mu_f z)$  is concave in  $z$ . Now if  $H_g^a(z)$  dominates over  $H_f^a(z)$  in the sense of second order stochastic dominance,  $H_f^a(z)$  is a mean preserving spread of  $H_g^a(z)$ . Then  $Es_g^a$  is greater than  $Es_f^a$ .<sup>19</sup> This completes the proof. ■

**Proposition 3** indicates if the distribution of  $\frac{f_a}{f}(y|a)$  is dominated in the sense of second order stochastic dominance by that of  $\frac{g_a}{g}(y|a)$ , information system  $f(y|a)$  is better (more efficient) than information system  $g(y|a)$  for a given action  $a$  since the former allows the principal to enforce such an action choice at less expected compensation cost.

The basic intuition for the proposition is as follows: Since every likelihood ratio distribution for a given action choice has the same mean value of zero, the second

---

<sup>19</sup>See Rothschild and Stiglitz (1970).

order stochastic dominance of  $\frac{q_a}{g}$  over  $\frac{f_a}{f}$  means the distribution of  $\frac{f_a}{f}$  is a mean preserving spread of that of  $\frac{q_a}{g}$ . When the likelihood ratio distribution under one information system is a mean preserving spread of that under the other information system, it provides more effective information to the principal by offering a more diverse set of data, in a probabilistic sense, from which he can draw inference about the agent's action choice.

The SOSD criterion is stronger than the variance criterion since the fact that the distribution of  $\frac{f_a}{f}$  is a mean preserving spread of that of  $\frac{q_a}{g}$  always means  $var(\frac{f_a}{f}) > var(\frac{q_a}{g})$  but the converse is not always true, and this SOSD criterion provides a partial ordering in the sense that two information systems to be compared may not have such a second order stochastic dominance relation.

It is natural to ask why we need a stronger criterion than the variance criterion in ranking information systems with the agent's general utility function. If the agent's utility function is a square root, by (10),  $w(z) = -(\lambda_f + \mu_f z)^2$  is quadratic in  $z$ . Thus, the Taylor expansion of  $w(z)$  around  $z = 0$  will end at the second moment of  $z$ .<sup>20</sup> This implies that the difference between  $Es_f^2$  and  $Es_g^2$  can be completely determined by the difference in the second moment of  $z$  since the first moment of  $z$  is zero for any information system. Thus, the second moment of  $z$ , which is equivalent to  $var(\frac{h_a}{h})$ , can alone be a measure in comparing the principal's expected compensation costs.

On the other hand, if the agent's utility function is not a square root, i.e., if  $w(z)$  is not quadratic in  $z$ ,  $var(\frac{h_a}{h})$  alone cannot serve as a relevant measure for the principal's expected compensation cost, since  $w''(z)$  is no longer independent of  $z$ . This reveals that the efficiency of an information system generally depends not only on the "amount of information", as measured by  $var(\frac{h_a}{h})$ , but on how it is used for contracting purposes, as affected by the agent's risk preferences. Thus, when

---

<sup>20</sup>The reason we expand  $w(z)$  around  $z = 0$  is that  $z$  under any information system has a mean value zero.

$w''(z)$  is not constant,  $\text{var}(\frac{h_a}{h})$  alone cannot capture the effect of the agent's risk preferences on the principal's expected compensation cost. In comparison, however, the SOSD criterion in **Proposition 3** is strong enough to dominate any of higher than second-order effects of  $w(z)$  on the expected compensation cost.

**Corollary 2:**

For any agent's preference satisfying **Assumption 1**, information system  $f(y|a)$  is 'more efficient' than  $g(y|a)$  if  $H_f^a(z)$  is *weakly dominated* by  $H_g^a(z)$  in the sense of second order stochastic dominance for all  $a \in (\underline{a}, \bar{a}]$ , and the (strong) second order stochastic dominance holds for at least one  $a \in (\underline{a}, \bar{a}]$ .

**Remark 1:**

A careful reader may have a question as to whether the converse of **Proposition 3** holds. The converse of **Proposition 3** is the statement that if information system  $f(y|a)$  is 'more efficient at  $a \neq \underline{a}$ ' than  $g(y|a)$  for any risk averse agent then the likelihood ratio distribution of that action choice under information system  $f(y|a)$ ,  $H_f^a(z)$ , is second order stochastically dominated by that under information system  $g(y|a)$ ,  $H_g^a(z)$ . Thus, if the converse is true, then any two information systems which can be ranked in their efficiencies in the agency model should have the second order stochastic dominance relation. As is well known in finance theory, the second order stochastic dominance criterion is an 'if and only if' condition for any risk averse economic agent to select a financial prospect. That is, any risk averse agent prefers one financial asset to the other if and only if the return distribution of the former dominates over that of the latter in the sense of second order stochastic dominance.<sup>21</sup> However, the converse of **Proposition 3** in the agency model has not yet been proved even though equation (9) which will be crucial for proving the converse of **Proposition 3** appears to be the same as the critical condition in the

---

<sup>21</sup>See Hadar and Russell (1969,1971), and Hanoch and Levy (1969).

analysis of a financial prospect. However,  $w(z)$  in equation (9) does not cover all the concave functions as the agent's utility function covers all increasing concave functions. Proving the converse or establishing a counter-example will be most helpful in clarifying the connections between the analysis here and Blackwell's theorem.

#### 4 Comparison with Holmstrom's Informativeness Criterion

In this section, we examine the relation between the SOSD criterion and the informativeness criterion in Holmstrom (1979). Using his notation, Holmstrom defines that the additional signal  $y$  is *informative* about the agent's action choice if and only if  $x$  is not a sufficient statistic for  $(x, y)$ , or equivalently it is false for some  $(x, y)$  and  $a$  that

$$h(x, y|a) = \phi(x, y)f(x|a),$$

where  $x$  denotes the commonly observed monetary outcome without cost.<sup>22</sup> He shows that the information system  $h(x, y|a)$  is 'more efficient' than  $f(x|a)$  if and only if the additional signal  $y$  is informative, i.e., it provides more information about the agent's action choice in addition to what is conveyed by  $x$ .

Kim and Suh (1990) show that Holmstrom's informativeness criterion can be equivalently represented by the variance criterion when the comparison is done for the inclusive information systems. This tells that the variance criterion as a measure of ranking information systems is better than Holmstrom's informativeness criterion in the sense that it can be applied not only to the comparison of inclusive information systems but also to the comparison of exclusive information systems with some

---

<sup>22</sup>This is the principal's cost minimizing version of Holmstrom's (1979) definition of informativeness with the agent's action choice fixed. In Holmstrom, the agent's action choice is not fixed for the comparison.

restrictions.<sup>23</sup> Now the following proposition shows that Holmstrom's informativeness criterion can also be equivalently represented by the SOSD criterion when the comparison is done for the inclusive information systems.

**Proposition 4:**

Let  $f(x|a)$  denote the information system which gives signal  $x$  while  $h(x, y|a)$  denotes the information system giving signal  $(x, y)$  where  $y$  represents the additional signal jointly observed. Then,  $H_h^a(z)$  is *weakly dominated* by  $H_f^a(z)$  for all  $a \in (\underline{a}, \bar{a}]$  in the sense of second order stochastic dominance, and the (strong) SOSD holds for at least one  $a \in (\underline{a}, \bar{a}]$  if and only if it is false that

$$h(x, y|a) = \phi(x, y)f(x|a) \text{ for almost every } (x, y) \text{ and } a \in (\underline{a}, \bar{a}].$$

As a preliminary to proving this proposition we start with the following lemma.

**Lemma 1:**

Two random variables  $X$  and  $Y$  are related as follows:

$$Y =_d X + Z$$

where “ $=_d$ ” means “has the same distribution as” and  $Z$  is a nontrivial random variable<sup>24</sup> with the property that

$$E(Z|X) = 0 \text{ for all } X$$

if and only if  $E(X) = E(Y)$  and the distribution of  $X$  (strongly) dominates over that of  $Y$  in the sense of second order stochastic dominance.

---

<sup>23</sup>It is already shown that the variance criterion ranks all the information systems when the agent's utility function is a square root. Kim and Suh (1990) also show that the variance criterion is an effective measure for the ranking of information systems when the set of information systems  $\Gamma$  is restricted to normal, lognormal, or double exponential (Laplace) family. For the details, see Kim and Suh (1990).

<sup>24</sup>Nontrivial random variable here means that it is false that  $Z = 0$  with probability 1.

**Proof of Lemma 1:**

See Rothschild and Stiglitz (1970).

**Proof of Proposition 4:**

(1) Sufficiency:

Denote  $h(x, y|a) = \phi(y|x, a)f(x|a)$ .

Then  $h_a(x, y|a) = \phi_a(y|x, a)f(x|a) + \phi(y|x, a)f_a(x|a)$

which yields

$$\frac{h_a(x, y|a)}{h(x, y|a)} = \frac{\phi_a(y|x, a)}{\phi(y|x, a)} + \frac{f_a(x|a)}{f(x|a)}$$

Let's denote  $Z_i = \frac{h_a}{h}$  where  $i = h, \phi, f$ . Then  $Z_h, Z_\phi$ , and  $Z_f$  are random variables which satisfy  $Z_h = Z_f + Z_\phi$ . If the aforementioned Holmstrom's informativeness condition holds, then

$$\frac{h_a(x, y|a)}{h(x, y|a)} \neq \frac{f_a(x|a)}{f(x|a)} \text{ for some } (x, y) \text{ and } a.$$

Thus,

$$\frac{\phi_a(y|x, a)}{\phi(y|x, a)} \neq 0 \text{ for some } (x, y) \text{ and } a.$$

So  $Z_\phi$  is a nontrivial random variable for such action  $a$ . Furthermore, it is obvious that  $E(Z_\phi|Z_f) = 0$  for all  $Z_f$  since

$$\int \frac{\phi_a(y|x, a)}{\phi(y|x, a)} \phi(y|x, a) dy = 0 \text{ for all } x \text{ and } a.$$

Since  $Z_h = Z_f + Z_\phi$  always means

$$Z_h =_d Z_f + Z_\phi,$$



by **Lemma 1** and by the fact that  $E(Z_h) = E(Z_f) = 0$  it follows that  $H_h^a(z)$  is weakly dominated by  $H_f^a(z)$  in the sense of second order stochastic dominance and the strong SOSD holds for some  $a$ .

(2) Necessity:

If Holmstrom's informativeness condition does not hold, then

$$\frac{h_a(x, y|a)}{h(x, y|a)} = \frac{f_a(x|a)}{f(x|a)} \text{ for almost every } (x, y) \text{ and } a.$$

Thus,

$$\frac{\phi_a(y|x, a)}{\phi(y|x, a)} = 0 \text{ for almost every } (x, y) \text{ and } a.$$

Then,  $Z_h = Z_f$  for all  $a \in (\underline{a}, \bar{a}]$ , which means  $Z_h =_d Z_f$  for all  $a \in (\underline{a}, \bar{a}]$ . This violates the requirement that the strong SOSD should hold for some  $a \in (\underline{a}, \bar{a}]$ .

■

Consequently, **Proposition 4** demonstrates that the aforementioned three criteria (Holmstrom's informativeness criterion, Variance criterion, SOSD criterion) become equivalent when the ranking of information systems is done for the inclusive information systems. However, since the SOSD criterion can be applied to the comparison of a broader set of information systems without any restriction either on the agent's preference or on the set of information systems to be ranked, it is much more effective as a measure in the respect that it includes Holmstrom's informativeness criterion and the variance criterion.

## 5 Comparison with Blackwell's Theorem

### Definition 6.<sup>25</sup>

Let  $Y_1$  and  $Y_2$  be random variables or experiments with probability density functions

---

<sup>25</sup>See DeGroot (1970)

$f(y_1|\omega)$  and  $g(y_2|\omega)$  respectively. The experiment  $Y_1$  is *sufficient* for the experiment  $Y_2$  if there exists a nonnegative function on the product space  $Y_1 \times Y_2$  for which the following three relations hold:

$$g(y_2|\omega) = \int h(y_1, y_2) f(y_1|\omega) dy_1 \quad \text{for all } y_2 \text{ and } \omega$$

$$\int h(y_1, y_2) dy_2 = 1 \quad \text{for all } y_1$$

and

$$0 < \int h(y_1, y_2) dy_1 < \infty \quad \text{for all } y_2$$

where  $\omega$  denotes the nature's unknown state.

Blackwell's theorem states that any decision maker prefers the information system  $f(y_1|\omega)$  to  $g(y_2|\omega)$  if and only if the experiment  $Y_1$  is sufficient for the experiment  $Y_2$ .<sup>26</sup>

In the agency model, the incentive scheme can be thought as the principal's decision function while the agent's action choice the unknown parameter. Thus Blackwell's theorem may suggest that ranking information systems in the agency model is formally equivalent to that in the Bayesian decision theory. To present the conclusion first, the sufficient part of Blackwell's theorem works in the agency model but whether the necessary part works is still questionable. Intuitively, the function  $h(y_1, y_2)$  generates a stochastic transformation from  $Y_1$  to  $Y_2$ .<sup>27</sup> Thus  $Y_2$  can be generated by means of randomization and it is obvious that the decision maker never chooses the experiment  $Y_2$  when  $Y_1$  is available because  $Y_2$  is simply the one obtained from adding white noise to  $Y_1$ .

### **Proposition 5:**

<sup>26</sup>See Blackwell (1951, 1953). For the simple proof of Blackwell's theorem, see Cremer (1982).

<sup>27</sup>Here we rule out nonstochastic single point transformation.

If information system  $f(y_1|a)$  is 'sufficient' for information system  $g(y_2|a)$ , then  $f(y_1|a)$  is 'more efficient at  $a$ ' than  $g(y_2|a)$  for all  $a \in (\underline{a}, \bar{a}]$ .<sup>28</sup>

**Proof:**

Let  $s_g^a(y_2)$  be the optimal incentive scheme in inducing the agent any given action  $a$  under the information system  $g(y_2|a)$ . To prove this proposition it will be enough to show that there always exists an incentive scheme under the information system  $f(y_1|a)$  which induces the agent the same action and also guarantees at least  $\bar{U}$  while it incurs less expected cost to the principal. Let's construct such an incentive scheme  $s_f(y_1)$  satisfying

$$u(s_f(y_1)) = \int h(y_1, y_2)u(s_g^a(y_2))dy_2 \quad (11)$$

Therefore,  $s_f(y_1)$  is a certainty equivalent wage scheme of  $s_g^a(y_2)$ . Then,

$$\begin{aligned} EU(s_g^a(y_2), a) &= \int u(s_g^a(y_2))g(y_2|a)dy_2 - v(a) \\ &= \int \int u(s_g^a(y_2))h(y_1, y_2)f(y_1|a)dy_1dy_2 - v(a) \\ &= \int u(s_f(y_1))f(y_1|a)dy_1 - v(a) \\ &= EU(s_f(y_1), a) \end{aligned}$$

Thus  $s_f(y_1)$  equally induces the agent  $a$  and also guarantees  $\bar{U}$  under the information system  $f(y_1|a)$ . But

$$\begin{aligned} Es_f &\equiv \int s_f(y_1)f(y_1|a)dy_1 \\ &< \int \int s_g^a(y_2)h(y_1, y_2)f(y_1|a)dy_1dy_2 \\ &= \int s_g^a(y_2)g(y_2|a)dy_2 = Es_g^a \end{aligned}$$

where the inequality comes from applying Jensen's inequality to (14).<sup>29</sup> Therefore.

<sup>28</sup>The statement that  $f(y_1|a)$  is 'more efficient at  $a$ ' than  $g(y_2|a)$  for all  $a \in (\underline{a}, \bar{a}]$  is stronger than that  $f(y_1|a)$  is 'more efficient' than  $g(y_2|a)$  because in the former statement the strict inequality between  $Es_f^a$  and  $Es_g^a$  holds for all  $a \in (\underline{a}, \bar{a}]$ .

<sup>29</sup>The strict inequality comes from the strict concavity of the agent's preference and the elimination of the nonstochastic transformation from  $h(y_1, y_2)$ .

the principal can reduce his expected cost by giving  $s_f(y_1)$  under information system  $f(y_1|a)$ . ■

**Proposition 5** illustrates that the sufficient part of Blackwell's theorem can be directly applied to the agency model in ranking information systems. It is intuitive that offering the risk averse agent the incentive scheme which is conditional on an additional lottery is never desirable rather than giving certainty equivalent income shown in (14).

However, about the converse little is known.<sup>30</sup> As a matter of fact, showing the converse of the **Proposition 5** is equivalent to proving that any two strictly ranked information systems *for all agent's action choices* in the agency model should satisfy Blackwell's sufficiency condition shown in **Definition 5**. But a simple characterization of this kind does not seem to exist. The game structure of the agency model is different from that of the Bayesian decision theory. In the agency model, the agent's action choice which is thought as a hidden parameter is not a pure random variable but the agent's (a player's) decision variable to be controlled by the principal. Thus the efficiency of an information system is not intrinsically related to the degree of statistical inference of the agent's hidden action but the degree of controlling the agent's action choice even though both are closely related. That is why the SOSD criterion is strongly based on the likelihood ratio of the information system, which captures the incentive provision for such controlling, rather than the likelihood function  $h(y|a)$  itself on which the Bayesian decision theory is based.

**Remark 2:**

Blackwell's notion of sufficiency in **Definition 5** is stronger than its general form in

---

<sup>30</sup>Gjesdal (1982) shows that the converse is not true in the agency model by giving a counterexample. That counterexample is composed of two information systems one is perfect and the other is a moving support. However, it is not appropriate in this paper since we have already ruled out both cases from  $\Gamma$ .

the sense that the nonstochastic single point transformation is ruled out in **Definition 5**. Also, it is a stronger criterion than the SOSD criterion since it gives the strict inequality between  $Es_f^a$  and  $Es_g^a$  for all  $a \in (\underline{a}, \bar{a}]$ .

**Remark 3:**

This proposition is a continuous action version which is parallel to the results obtained in Grossman and Hart (1983) and Gjesdal (1982). This is strongly based on **Assumption 1**. Sometimes a randomized incentive scheme can be better in the absence of **Assumption 1**.<sup>31</sup>

## 6 Welfare Implication

It has been recognized that the presence of moral-hazard incurs social welfare loss. The first best solution in contracting between the risk neutral principal and the risk averse agent is providing a constant wage scheme and enforcing the first best action.<sup>32</sup> This constant wage scheme also implies the constant risk sharing rule between the principal and the agent in equation (1). However, the moral-hazard problem induces the principal to deviate from the first best risk sharing rule for incentive provision. This obviously reduces the principal's expected revenue. Furthermore, as pointed out frequently, the moral-hazard problem may result in the deviation from the first best action since it may be too costly for the principal to induce the agent to take such an action choice. Therefore, the sources of welfare loss with moral-hazard have been thought as the deviations from the first best risk sharing rule and action choice.

However, this argument is sufficient only when the information system is socially given to the principal. When the principal is facing several alternatives as infor-

---

<sup>31</sup>See Gjesdal (1982).

<sup>32</sup>The first best action occurs where the expected marginal productivity equals the agent's marginal disutility of effort.

mation systems to select, there may be an additional source of welfare loss. For the ease of exposition, let us suppose that the principal has two alternatives as information systems  $f$  and  $g$  and the only observable is the outcome  $x$ , that is,  $y = x$ . Then the principal's two alternatives can be thought as two production projects to be selected. Now relax **Assumption 5** and suppose that project  $g$  gives more expected output than  $f$  for all action choice while project  $f$  provides more precise information satisfying **Proposition 3**. If the benefit from the precise information more than offsets the benefit from the high productivity, the principal will choose project  $f$  instead of  $g$  which should be chosen without moral-hazard. Observing the deviation associated with  $f$  thus understates the welfare loss since it ignores the switch to a less productive information system. That is, the deviation from the first best project choice is a third source of welfare loss.

## 7 Conclusion

The main pursuit of this paper is to find a criterion which makes it possible to rank a broader set of information systems in an agency model. The SOSD criterion is derived and its economic meaning is discussed. This SOSD criterion applies to a broader set of information systems in the sense that the information systems to be ranked do not need to have Holmstrom's sufficient statistic relationship. Of course, this criterion still provides a partial ordering because the information systems to be compared should have the second order stochastic dominance relationship. The economic intuition is that with the agent's general utility function the efficiency of an information system is not only based on the physical amount of information system about the agent's hidden action choice but on the effective amount of information that is related to the agent's risk aversion.

The SOSD criterion for the ranking of information systems is specific to the principal-agent model. Blackwell's theorem, on the other hand, states that all decision-makers in all decision situations will prefer one information system to another *if and only if* the former is sufficient for the latter in terms of his notion of sufficiency. However, the game structure of the agency model is different from that of classical decision theory. In the agency model, the agent's action choice to be statistically inferred by the principal is not a pure random variable, that is, not the nature's decision variable but the agent's choice variable to be controlled by the principal through the incentive scheme while in decision theory the variable to be statistically inferred is purely random. We show that the sufficient part of Blackwell's theorem can be directly applied to the ranking of information systems in the agency model. However, it is still questionable whether the necessary part of Blackwell's theorem works in the agency model.

Another question is whether parallel results can be found in general contract theo-

ries that arise from more general informational asymmetries including not just hidden action but hidden knowledge (“adverse selection”). In hidden knowledge models, the principal’s problem of designing a contract is to give the agent an incentive to reveal his hidden knowledge. So the game structure is similar to that of agency model and the answer is positive. However, as usually assumed in this kind of models, if there is no common observable which is correlated with the *risk neutral* agent’s hidden variable (e.g., the agent’s hidden quality or other private informations) but the principal has only a prior belief about such hidden variable, then the difference in the preciseness of the principal’s prior will not make any difference in the principal’s welfare loss from the first best even though the incentive scheme will be sensitively related to the given prior.<sup>33</sup> This means it is meaningless to investigate the efficiency of an information system according to the preciseness of the principal’s prior because the principal’s given prior cannot play a role of ‘information’ about the agent’s hidden knowledge in designing an incentive scheme. However, as in the agency model, suppose that there is a common observable which is imperfectly correlated with the agent’s hidden variable. Then such observable can play an ‘informational’ role in designing an incentive scheme and it will be meaningful to check the efficiency of an information system according to the preciseness of the observable, that is, how closely the observable is correlated with the agent’s hidden variable.

An instant and fruitful extension will come from proving the converse of **Proposition 3** or providing a counter-example. As stated earlier, such analysis will make clear the relationship between the Bayesian decision theory and the agency model. Other possible extensions will come from relaxing the agent’s additively separable preferences on wealth and effort, or relaxing the assumption that the first order ap-

---

<sup>33</sup>The priors which is different in their preciseness here means different priors but with the same expected value about the agent’s hidden variable. And the ‘first best’ will be achieved when the principal can exactly observe the hidden variable.



proach is valid.<sup>34</sup>

We conclude with a couple of simple applications of the central results. One possible application is concerned with the principal's optimal investment into conditional monitoring. In the conditional monitoring setting, the principal chooses the amount of monitoring investment after he observes the outcome, and it is generally assumed that more investment into monitoring enhances the precision of an additional signal generated by the investment. Thus, varying the monitoring investment by the principal is equivalent to manipulating the efficiency of an information system. Moreover his optimal monitoring investment strategy is technically similar to his choosing information system in its efficiency. The main difference is that the principal manipulates the information system after he has observed the outcome. Thus, it is more closely concerned with the efficiency of the additional signal generated by such investment.<sup>35</sup>

Another example is related to the role of collateral as an incentive device in the credit market. It has been argued that, under limited liability contracts, a risk neutral borrower has an incentive to choose a project that is undesirably risky to a lender.<sup>36</sup> This conflicting incentive of the borrower cannot be reconciled unless the collateral fully covers his liabilities. However, in a more realistic situation where the borrower chooses a project and then endows it to a risk averse worker who chooses his level of hidden effort, partial collateral can work for reconciling the borrower's incentive in choosing a project with the lender's. This is a two-layer moral hazard problem. In the first layer, the borrower, as an agent, has an incentive to choose a risky project given the limited liability contract. But, in the second layer, he, as a principal, has an incentive to endow a riskless project and so makes it less costly

---

<sup>34</sup>It can be shown that the main result (Proposition 3) still holds in the case of discrete action choice under certain assumptions. Those assumptions are closely related to the validity of the first order approach.

<sup>35</sup>The issue of optimal monitoring investment is investigated in Kim and Suh (1989).

<sup>36</sup>For the details of this issue, see Stiglitz and Weiss (1981).

to induce high effort by the worker. Thus it may be possible to induce the riskless project with only partial collateral.

## REFERENCES

- Blackwell, D. 1951 "Comparison of Experiments." in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press.
- Blackwell, D. 1953 "Equivalent Comparison of Experiments.", *Annals of Mathematics and Statistics* 24: 265-272.
- Cremer, J. 1982, "A Simple Proof of Blackwell's 'Comparison of Experiments' Theorem." *Journal of Economic Theory* 27: 439-443.
- De Groot, M. 1970, "Optimal Statistical Decisions.", New York, McGraw-Hill Book Company.
- Freeman, H. 1963. "Introduction to Statistical Inference.", Addison-Wesley.
- Gjesdal, F. 1982 "Information and Incentives: The Agency Information Problem." *Review of Economic Studies* 373-390.
- Grossman, S., and O. Hart 1983. "An Analysis of the Principal-Agent Problem." *Econometrica* 51:7-45.
- Hadar, J. and Russell, W.R. 1969. "Rules for Ordering Uncertain Prospects." *American Economic Review* 59: 25-34.
- Hadar, J. and Russell, W.R. 1971. "Stochastic Dominance and Diversification." *Journal of Economic Theory* 3: 288-305
- Hanoch, G. and Levy, H. 1969. "The efficiency Analysis of Choices Involving Risk." *Review of Economic Studies* 36: 335-346.
- Harris, M., and Raviv, A. 1979. "Optimal Incentive Contracts with Imperfect Information." *Journal of Economic Theory* 20:231-259.
- Hart, O., and Holmstrom, B. 1985. "Theory of Contracts." presented in the 5th World Congress of Econometric Society.
- Holmstrom, B. 1979. "Moral Hazard and Observability." *Bell Journal of Economics*

10: 74-91.

Holmstrom, B. 1982. "Moral Hazard in Teams" *Bell Journal of Economics* 13:324-340.

Jewitt, I. 1988. "Justifying the First Order Approach to Principal-Agent Problems." *Econometrica* 56:1177-1190.

Kim, S. K., and Suh, Y. S. 1989. "Conditional Monitoring in Agency Model" *UCLA unpublished manuscript*

Kim, S. K. and Suh, Y. S. 1990. "Value of Information System and Income Smoothing." *UCLA unpublished manuscript*

Marschak, J. and Miyazawa, K. 1968. "Economic Comparability of Information systems" *International Economic Review* 9: 137-174.

Mirrlees, J. 1974. "Notes on Welfare Economics, Information and Uncertainty." In M. Balch, D. McFadden, and S. Wu (eds.), *Essays in Economic Behavior Under Uncertainty*, pp. 243-258. Amsterdam: North-Holland.

Rogerson, W. 1985. "The First Order Approach to Principal-Agent Problem." *Econometrica* 53:1357-1368.

Ross, S. 1973. "The Economic Theory of Agency: The Principal's Problem." *American Economic Review* 63: 134-139.

Rothschild, M., and Stiglitz, J. 1970. "Increasing Risk I: A Definition." *Journal of Economic Theory* 2:225-243.

Shavell, S. 1979. "Risk Sharing and Incentives in the Principal and Agent Relationship." *Bell Journal of Economics* 10:55-73.

Stiglitz, J., and A. Weiss 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71(3):393-410.