An Agent's Preferences and the principal's Incentive Cost in the Agency Problem

Jin Yong Jung*

January, 2017

Abstract

We provide the general results about the question of what effects the change of an agent's preferences over income and effort has on a risk-neutral principal's incentive cost. We argue that there are two factors affecting the principal's incentive cost. One is the agent's risk aversion and the other is his incentive sensitivity. We show that the increase in the agent's risk aversion or the decrease in his incentive sensitivity leads to the increase in the principal's incentive cost. And, we show that it is possible that the principal prefers the more risk averse agent.

1 Introduction

In the principal-agent problem, if an agent is risk-neutral, the risk-neutral principal under full information (i.e., when she can observe the agent's action choice) could pay to the agent as much as the expected compensation cost under asymmetric information (i.e., when she cannot observe it). However, if the agent is risk-averse, the expected compensation cost under

*Seoul National University, Department of Economics.

asymmetric information is greater than that under full information. These mean that if the agent is risk-neutral, the principal's incentive cost (i.e., the difference of the compensation costs between the asymmetric information case and the full information case) is zero, while it is positive if the agent is risk-averse. From this observation, one has believed that the increase in the agent's degree of risk aversion may cause the increase in the principal's incentive cost.

Grossman and Hart (1983) initially showed the relationship between the agent's risk aversion and the principal's incentive cost in a somewhat restrictive agency problem.[1] Especially, in the case that the agent has a multiplicatively separable CARA utility function (i.e., $U(s,a) = -e^{-k(s-a)}$), his reservation utility level is $\overline{U} = -e^{-k\alpha}$, and there are only two possible outcomes, they verified that, as the agent becomes more risk-averse (i.e., when the degree of absolute risk aversion, k, increases), the principal's incentive cost increases.[2] Later, Chade and Serio (2002) generalized the result of Grossman and Hart (1983) by extending the number of outcomes to n.

However, Chade and Serio (2002) merely double check the fact that the agent's risk aversion and the principal's incentive cost have a positive relationship only when his utility is CARA. Moreover, like Grossman and Hart (1983), they did not give any logical explanation about why the principal's incentive cost increases as the agent becomes more risk-averse. Thus, even the relationship between the agent's risk aversion and the principal's incentive cost has not yet been proved completely.

The purpose of this paper is to provide the more general results on what effects the change of the agent's preferences over income and effort has on the principal's incentive cost. To this end, we consider the standard principal-agent model where a risk-neutral principal

[1]See the Proposition 17 in Grossman and Hart (1983).

[2]They showed that the increase of k causes the increase in the principal's expected compensation cost under asymmetric information. However, in their model, since the principal's compensation cost under full information is equal to $\alpha + a$ which is independent of k, what they actually showed is that as k increases, the principal's incentive cost increases.

employs one of two agents with heterogeneous preferences.

We argue that two factors with respect to the agent's preferences over income and effort affect the principal's incentive cost: one is his risk-aversion, and the other is his incentive sensitivity. More precisely, the agent's risk aversion is represented by his degree of absolute risk aversion around his fixed wage level and his incentive sensitivity is represented by his marginal rate of substitution between income and effort, which is equal to his marginal utility of income divided by his marginal cost of effort in the case that the agent has an additively separable utility function. We show not only that the agent's incentive sensitivity negatively affects the principal's incentive cost, but also that the agent's risk aversion positively affects the principal's incentive cost.

The positive relationship between the agent's risk aversion and the principal's incentive cost shares the common intuition with the well-known Pratt's result about the relationship between a consumer's risk aversion and his risk premium in the asset market. This intuition is well revealed only when one understands the fact that the incentive cost on the principal's side is regarded as the risk premium on the agent's side. Note that the risk averse agent is compensated from the risk-neutral principal with a risk-free wage (i.e., a fixed wage) when his action is observable, but with a risky wage (i.e., an incentive wage) when his action is unobservable. When compensated with the risky wage, the risk-averse agent will claim higher expected wage level than the risk-free wage level. This difference between the expected risky wage level and the risk-free wage level is interpreted by the risk premium that the agent demands, and thus becomes the incentive cost on the principal's side. Thus, the relationship between the agent's risk aversion and the principal's incentive cost is identical to the relationship between his risk aversion and risk premium. As a result, as the agent becomes more risk averse, the principal's incentive cost increases, because he will claim the higher risk premium.

We also argue that the agent's incentive sensitivity and the principal's incentive cost have a negative relationship. Note that the agent's incentive sensitivity is represented by the marginal rate of substitution between income and effort. Then, an improvement of the agent's incentive sensitivity means that his response to an incentive wage which was previously provided by the principal becomes more sensitive. Thus, when the agent's incentive sensitivity is improved, he will work harder under the given incentive wage contract. At this time, the principal who wants to induce the same effort level can design the cheaper wage contract for the agent, which implies that the principal's incentive cost decreases.

In fact, Chade and Serio (2002) failed to figure out the effect of the agent's incentive sensitivity on the principal's incentive cost. It comes from special characteristics of the agent's utility function and reservation utility that they considered in their model. We need to notice that, in the case that the agent with reservation utility $\overline{U} = -e^{-k\alpha}$ has utility function $U(s, a) = -e^{-k(s-a)}$, his marginal rate of substitution between income and effort is always the one even if the value of k changes. This means that the agent's incentive sensitivity is independent of the value of k. Thus, their model has the limitation that the change of k brings about the change of the agent's risk aversion, but never changes his incentive sensitivity. As a result, it is an example to show the relationship between the agent's risk aversion and the principal's incentive cost only when his incentive sensitivity is unchanged.

Note that, since the agent's incentive sensitivity is represented by the ratio of his marginal utility of income and marginal cost of effort, it is positively affected by his marginal utility, but negatively by his marginal cost. Now, suppose that the marginal cost is unchanged. Then, the agent's incentive sensitivity is directly linked to the magnitude of his marginal utility. In this case, the increase in the agent's marginal utility implying the improvement of his incentive sensitivity will reduce the principal's incentive cost. This indicates that not only the agent's risk aversion but also the agents' marginal utility is an important factor for the principal.

Our argument about the negative relationship between the agent's marginal utility and the principals' incentive cost is beyond the existing common knowledge in the expected utility theory. Note that an individual's marginal utility itself is not important at all in the financial decision theory.[3] However, our argument that the agent's marginal utility negatively affects the principal's incentive cost is natural to the principal who should give the agent an incentive to work as hard as ever.

The similarities and differences between our results in the agency problem and the existing Pratt's result in the consumer problem come from the fact that the decision maker (principal) in the agency problem has to consider one more issue than a decision maker (consumer) in the consumer problem. In the consumer's problem, only the issue of consumer's market participation is considered, whereas, in the agency problem, not only the issue of the agent's contract participation but also the issue of his incentive provision must be considered. Then, the participation issue which is commonly considered in both problems makes the same result that the increase of the degree of risk aversion causes the higher risk premium or incentive cost. However, the incentive provision issue which is considered only in the agency problem makes the difference that the agent's marginal utility itself is also important in the agency problem.

The paper is organized as follows: In Section 2, we formulate the basic principal-agent framework. In section 3, our results on the effects of the agent's incentive sensitivity and his risk aversion on the principal's incentive cost are provided and discussed. Concluding remarks are offered in Section 4, and all the formal proofs are relegated to Appendix.

[3] This is the reason to be said that an individual's utility function is unique up to any affine transformation.

2 The basic model

We consider a one-period standard principal-agent model in which risk-averse agent $i = 1, 2$ works for a risk-neutral principal. In the beginning of the period, each agent inputs his action (or effort) $a \in A \equiv [0, \overline{a}]$ into a stochastic production technology, and then in the end of the period, commonly observable output $x \in X \subseteq \mathbb{R}$ is realized.

Since output x is stochastically correlated with each agent's action a, a cumulative distribution function of x conditional on a is denoted by $F(x|a)$ and the corresponding probability density(mass) function is denoted by $f(x|a)$ where $F(x|a)$ and $f(x|a)$ are at least twice differentiable with respect to a. We assume that density function $f(x|a)$ satisfies the MLRP (monotone likelihood ratio property) and its likelihood ratio $\frac{f_a(x|a)}{f(x|a)}$ is lower bounded.[4] Also, we assume that the support X is independent of a.[5] Finally, we assume that $f(x|a)$ satisfies the CDFCL (Convexity of the Distribution Function Condition for the Likelihood-ratio) for the validity of the first-order approach.[6]

Agent i's utility function is given by an additively separable form: $U_i(s, a) = u_i(s) - c_i(a)$. $u_i(s)$ denotes agent i's utility from monetary payoff s with $u_i' > 0$ and $u_i'' < 0$, and so he is risk-averse. $u_1(s)$ and $u_2(s)$ are defined on the same domain $(\underline{d}, \overline{d})$ where $\lim_{s \downarrow \underline{d}} u_i'(s) = \infty$ and $\lim_{s \uparrow \overline{d}} u_i'(s) = 0$ for all $i = 1, 2$. $c_i(a)$ denotes agent i's disutility from action with

[4] If the likelihood ratio $\frac{f_a(x|a)}{f(x|a)}$ may go to $-\infty$ on the support X, the existence of the optimal contracts is not guaranteed, which is called as the Mirrlees' unpleasant case.

[5] In the agency problem, the case that the support X is dependent on a gives the trivial result that the second best optimal contract is same with first best optimal contract, which is called as the moving support problem. For a detailed explanation, see Gjesdal (1982).

[6] The first-order approach is the method of replacing the original argmax incentive compatibility constraint with the first order condition of the agent's expected utility function with respect to a. The most generalized results for justifying the first-order approach when the principal is risk-neutral are Jung and Kim's two sets of conditions: the CDFCL, and a set of two conditions of Proposition 7 in Jung and Kim (2015). In this paper, we assume the CDFCL for $f(x|a)$ for a general analysis. For a detailed explanation, see Jung and Kim (2015).

$c_i'(a) > 0$ and $c_i''(a) \geq 0$, and so he is work-averse. $\overline{U}_i$ denotes agent i's reservation utility. In order to eliminate trivial cases, it is assumed that $u_i(\underline{d}) \leq c_i(0) + \overline{U}_i$ for all $i = 1, 2$.

Fix $a > 0$. As is well known, the risk-neutral principal should compensate risk-averse agent i with fixed salary $s_i^f \equiv u_i^{-1}(c_i(a)+\overline{U}_i)$ under full information where she can observe his effort a. On the other hand, under asymmetric information where his effort a is unobservable to her, the principal should design other wage contract depending on x for him, which is denoted by $w_i(x) + s_i^f$. Then, in order to induce agent i to voluntarily choose a given action $a > 0$ at minimum cost, the principal should solve the following cost minimization problem:

$$\begin{aligned} \min_{w_i(x)} \quad & \int w_i(x) f(x|a) dx \\ s.t. \quad i) & \int u_i(w_i(x) + s_i^f) f(x|a) dx - c_i(a) \geq \overline{U}_i, \\ ii) & \int u_i(w_i(x) + s_i^f) f_a(x|a) dx \geq c_i'(a), \end{aligned}$$

where the first and the second constraints are called the participation and the (doubly relaxed) incentive compatibility constraints for agent i, respectively.

Let $w_i^a(x)$ be the optimal incentive contract for agent i. Then, $w_i^a(x)$ should satisfy

$$\frac{1}{u_i'(w_i^a(x) + s_i^f)} = \lambda_i + \mu_i \frac{f_a(x|a)}{f(x|a)}, \tag{1}$$

for almost every $x \in X$, where λ_i and μ_i are the Lagrangian multipliers of the participation and the incentive compatibility constraints, respectively.[7] Since the signs of λ_i and μ_i are positive by Lemma 1 in Jewitt (1988), both the participation and the incentive constraints should be binding at the optimum.[8]

[7]For the existence of the optimal contract $w_i^a(x)$, it must be required that $\lambda_i + \mu_i \frac{f_a(x|a)}{f(x|a)} > 0$ for almost every $x \in X$. Moroni and Swinkels (2014) showed that some extra conditions for the agent's utility or cost function should be needed in order to satisfy that inequality for all x. For a detailed explanation, see Moroni and Swinkels (2014).

[8]The Lemma 1 in Jewitt (1988) is actually proved in the case that the relaxed incentive compatibility constraint is considered. However, even in the case that the "doubly relaxed" incentive constraint is considered, his proof works well.

For analytic simplicity, we let $\hat{u}_i(s) \equiv \frac{1}{c_i'(a)}[u_i(s+s_i^f) - u_i(s_i^f)]$, for all $i = 1, 2$. Note that $\hat{u}_i(s)$ is an affine transformation of $u_i(s + s_i^f)$ which is obtained by parallel shifting $u_i(s)$ in the direction of x-axis by $-s_i^f$.[9] Note that $\hat{u}_i(s)$ goes through the origin at which $\hat{u}_1(s)$ and $\hat{u}_2(s)$ always crosses. Then, we have

$$\begin{aligned} E[\hat{u}_i(w_i^a(x))] &\equiv \int \hat{u}_i(w_i^a(x)) f(x|a) dx \\ &= \frac{1}{c_i'(a)} \left[\int u_i(w_i^a(x) + s_i^f) f(x|a) dx - u_i(s_i^f) \right] = 0, \end{aligned} \tag{2}$$

and

$$\begin{aligned} \int \hat{u}_i(w_i^a(x)) f_a(x|a) dx &= \frac{1}{c_i'(a)} \int [u_i(w_i^a(x) + s_i^f) - u_i(s_i^f)] f_a(x|a) dx \\ &= \frac{1}{c_i'(a)} \int u_i(w_i^a(x) + s_i^f) f_a(x|a) dx = 1, \end{aligned} \tag{3}$$

where the second equality holds by definition and by the fact that $\int f_a(x|a)dx = 0$.

In the paper, $E[w_i^a(x)] \equiv \int w_i^a(x) f(x|a) dx$ is interpreted as the principal's incentive cost from agent i. Note that $E[w_i^a(x)] + s_i^f$ is the principal's expected compensation cost in the asymmetric information case. On the other hands, as explained earlier, s_i^f is the principal's compensation cost in the full information case. Then, $E[w_i^a(x)]$ is the principal's net cost that she should additionally pay to agent i for incentive provision at a minimum.[10]

For any given $a > 0$, $E[w_i^a(x)] \geq 0$ for any distribution $f(x|a)$. Note that, from the definition of $\hat{u}_i(s)$,

$$\hat{u}_i(0) = 0, \tag{4}$$

for all $i = 1, 2$. Then, by combining (2) and (4) we have $E[\hat{u}_i(w_i^a(x))] = \hat{u}_i(0)$, to which applying Jensen's inequality gives $\hat{u}_i(E[w_i^a(x)]) \geq \hat{u}_i(0)$, or $E[w_i^a(x)] \geq 0$. From this, one can see that the incentive cost from agent i is zero (i.e., $E[w_i^a(x)] = 0$) when he is risk-neutral

[9] We allow the case that $s_1^f \neq s_2^f$.

[10] In fact, the incentive cost means the principal's loss from being unable to observe his action choice, which is used in Grossman and Hart (1983) and in Chade and Serio (2003).

(i.e., $u_i''(s) = 0$), while his incentive cost is positive (i.e., $E[w_i^a(x)] > 0$) when he is risk-averse (i.e., $u_i''(s) < 0$). These observations have raised the question: As an agent's degree of risk aversion increases, will his incentive cost increase? This is one part of the main questions of the paper. The purpose of this paper is to show what effects the change of an agent's preferences over income and effort has on the principal's incentive cost.

3 An Agent's Preferences and his Incentive Cost

The question about the relationship between an agent's risk-aversion and the principal's incentive cost has been initially raised by Grossman and Hart (1983). Grossman and Hart showed that they have a positive relationship in a somewhat restrictive model where the agent has multiplicatively separable CARA utility, i.e., $U(s, a) = -e^{-k(s-a)}$, his reservation utility is $\overline{U} = -e^{-k\alpha}$, the action set A is finite, and the number of outcomes of output is equal to two.[11] Later, Chade and Serio (2002) generalized the Grossman and Hart's (1983) result by extending the number of outcomes to n. They showed that, as the agent becomes more risk-averse in the sense of constant absolute risk aversion (i.e., as k increases), the principal's incentive cost also increases.

However, Chade and Serio (2002) gave just one example to reveal a positive relationship between the agent's risk aversion and the principal's incentive cost.[12] Furthermore, they did not provide any intuitive explanation about why the increase in the agent's degree of risk aversion brings about the increase in the principal's incentive cost. Thus, the question about that relationship has still been left veiled.

We argue that there are two factors with respect to the agent's preferences, which affects the principal's incentive cost: One is the agent's risk aversion which is measured as the

[11] See Proposition 16 in Grossman and Hart (1983).

[12] Actually, they do not give any answer about that relationship when an agent has a general utility function, not CARA.

degree of absolute risk aversion around his fixed salary (i.e., $-\frac{u_i''(s+s_i^f)}{u_i'(s+s_i^f)}$), and the other is his incentive sensitivity which is measured as the marginal utility of income over the marginal disutility of effort (i.e., $\frac{u_i'(s_i^f)}{c_i'(a)}$). Note that the agent's incentive sensitivity means his marginal rate of substitution between income and effort i.e., $\frac{da}{ds}\big|_{U^i(s,a)=\overline{U}_i} = \frac{u_i'(s_i^f)}{c_i'(a)}$. Thus, the agent's incentive sensitivity indicates how harder he will work when his wage increases by 1 unit at s_i^f.

Our argument will be discussed through the three steps. The first step is to analyze the relationship between the agent's incentive sensitivity and the principal's incentive cost, keeping his degree of risk aversion fixed. Later, we will analyze the relationship between the agent's degree of risk aversion and the principal's incentive cost, keeping his incentive sensitivity fixed. Finally, the combined effect of the agent's risk aversion and incentive sensitivity on the principal's incentive cost will be analyzed.

We start with the following lemma.

Lemma 1: *If i)* $\hat{u}_1'(s) \geq \hat{u}_2'(s)$ *for all* s *and ii)* $E[\hat{u}_1(w_2^a(x))] \geq 0$, *then* $E[w_1^a(x)] \leq E[w_2^a(x)]$.

Suppose that agent i has the parallel-shifted utility function $\hat{u}_i(s) = \frac{1}{c_i'(a)}[u_i(s+s_i^f) - u_i(s_i^f)]$. Lemma 1 says that, if the marginal utility of agent 1 is not less than that of agent 2, then agent 1 has not greater incentive cost than agent 2, as long as the participation constraint for agent 1 is satisfied when the optimal incentive wage for agent 2, $w_2^a(x)$, is offered to him. As seen in the proof of Lemma 1, the reason is that, since $w_2^a(x)$ satisfies the participation constraint for agent 1 by condition ii) in Lemma 1, and since $w_2^a(x)$ satisfies the doubly relaxed incentive constraint for agent 1 by condition i) in Lemma 1, $w_2^a(x)$ belongs to the principal's opportunity set for $w_1(x)$ satisfying the participation and the incentive constraints for agent 1. This means that the principal can design the cheaper incentive wage contract for agent 1 than $w_2^a(x)$ in order to induce the same action level a.

By using Lemma 1, we analyze the pure effect of the agent's marginal utility on the principal's incentive cost.

Proposition 1: *If i)* $\frac{u_1'(s_1^f)}{c_1'(a)} \geq \frac{u_2'(s_2^f)}{c_2'(a)}$*, and ii)* $-\frac{u_1''(s+s_1^f)}{u_1'(s+s_1^f)} = -\frac{u_2''(s+s_2^f)}{u_2'(s+s_2^f)}$ *for all* s*, then* $E[w_1^a(x)] \leq E[w_2^a(x)]$.

Note that applying the definition of $\hat{u}_i(s)$ to conditions i) and ii) in the above proposition gives $\hat{u}_1'(0) \geq \hat{u}_2'(0)$ and $-\frac{\hat{u}_1''(s)}{\hat{u}_1'(s)} = -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}$ for all s, respectively. Again, suppose that agent i has the parallel-shifted utility $\hat{u}_i(s)$. Then, this proposition shows that, if the two agents have the same degree of absolute risk aversion, agent 1 with the greater marginal utility at 0 has the lower incentive cost than agent 2 with the less marginal utility at 0.

As shown in the proof of Proposition 1, note that, when agents 1 and 2 have the same degree of absolute risk aversion, the parallel-shifted utility function for agent 1, $\hat{u}_1(s)$, is represented by a linear transformation of that of agent 2, $\hat{u}_2(s)$, such as $\hat{u}_1(s) = \alpha\hat{u}_2(s)$ where $\alpha = \frac{\hat{u}_1'(0)}{\hat{u}_2'(0)}$. Then, if agent 1's marginal utility at 0 is not less than agent 2's marginal utility at 0, α is not less than 1, which implies that the marginal utility of agent 1 is not less than that of agent 2 for all s (i.e., $\hat{u}_1'(s) \geq \hat{u}_2'(s)$ for all s). Furthermore, when $w_2^a(x)$ is offered to agent 1, his expected utility is $E[\hat{u}_1(w_2^a(x))] = \alpha E[\hat{u}_2(w_2^a(x))] = 0$, which means that $w_2^a(x)$ satisfies the participation constraint for agent 1. Thus, by using Lemma 1, we have that the principal's incentive cost from agent 1 is not greater than that from agent 2.

As explained earlier, term $\frac{u_i'(s_i^f)}{c_i'(a)}$ in condition i) of Proposition 1 means agent i's incentive sensitivity. Then, condition i) in Proposition 1 indicates that agent 1 has the more sensitive response to an incentive wage commonly provided by the principal than agent 2. Thus, what Proposition 1 actually shows is that the agent's incentive sensitivity has a negative effect on the principal's incentive cost. As seen in Proposition 1, when agents 1 and 2 have the same degree of absolute risk aversion around their own fixed salaries (i.e., $-\frac{u_1''(s+s_1^f)}{u_1'(s+s_1^f)} = -\frac{u_2''(s+s_2^f)}{u_2'(s+s_2^f)}$ for all s), if agent 1's incentive sensitivity is not lower than agent 2's (i.e., $\frac{u_1'(s_1^f)}{c_1'(a)} \geq \frac{u_2'(s_2^f)}{c_2'(a)}$), then the principal's incentive cost from agent 1 is not greater than that from agent 2.

Therefore, when the agent's incentive sensitivity increases with his risk aversion unchanged, the principal's incentive cost decreases.

The reason why the agent's incentive sensitivity negatively affects the principal's incentive cost is that it is related with his response to an incentive wage which is provided around his fixed salary s_i^f. Note that the optimal incentive wage for agent i, $w_i^a(x)$, always goes through zero (i.e., x-axis) from the fact that $E[u_i(w_i^a(x)+s_i^f)] = u_i(s_i^f)$. Let us consider an arbitrary incentive wage $w(x)$ which is provided around 0. And, suppose that, when $w(x)+s_i^f$ is offered to agent i for all i, and when the two agents make a choice of the same action a, their expected utility levels are equal.[13] At this time, if agent 1's incentive sensitivity is greater than agent 2's, agent 1 will work harder than agent 2, because agent 1 has the more sensitive response to the given incentive wage $w(x)$ than agent 2. Thus, the principal who wants to induce the same action a can design the cheaper incentive wage contract for agent 1 than that for agent 2. As a result, since an agent's incentive sensitivity indicates his response to an incentive wage provided around his fixed salary, it has a negative effect on the principal's incentive cost.

Now, we analyze the pure effect of the agent's risk aversion on the principal's incentive cost. It is possible by keeping his incentive sensitivity fixed.

Proposition 2: *If i)* $\frac{u_1'(s_1^f)}{c_1'(a)} = \frac{u_2'(s_2^f)}{c_2'(a)}$, *and ii)* $-\frac{u_1''(s+s_1^f)}{u_1'(s+s_1^f)} \leq -\frac{u_2''(s+s_2^f)}{u_2'(s+s_2^f)}$ *for all* s, *then* $E[w_1^a(x)] \leq E[w_2^a(x)]$.

The above proposition shows a positive relationship between the agent's risk aversion and the principal's incentive cost. When we consider agent i with parallel-shifted utility $\hat{u}_i(s)$, Proposition 2 says that, if agent 2 is more risk-averse than agent 1 under the degree of absolute risk aversion (i.e., $-\frac{\hat{u}_1''(s)}{\hat{u}_1'(s)} \leq -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}$ for all s), then the principal's incentive cost

[13] As seen in the proof of Proposition 1, an example of such an incentive wage is the optimal incentive wage for agent 2, $w_2^a(x)$.

from agent 2 is higher than that of agent 1, as long as their marginal utilities at 0 are equal (i.e., $\hat{u}_1'(0) = \hat{u}_2'(0)$). Thus, this proposition shows that the more risk averse agent causes the higher incentive cost only when his incentive sensitivity is kept unchanged.

In order to understand the result of Proposition 2 more precisely, let us consider the two-outcome case where $x \in \{x_1, x_2\}$. Let $p_j = f(x_j|a)$ and $q_j = \frac{f_a(x_j|a)}{f(x_j|a)}$ for all $j = 1, 2$. Note that $\sum_j f(x_j|a) = p_1 + p_2 = 1$ and $\sum_j f_a(x_j|a) = q_1 p_1 + q_2 p_2 = 0$. And, note that $q_1 < 0 < q_2$ by the MLRP for $f(x|a)$. For agent $i = 1, 2$, since the participation and incentive constraints are binding at the optimum, equations (2) and (3) are equal to $\hat{u}_i(w_i^a(x_1))p_1 + \hat{u}_i(w_i^a(x_2))p_2 = 0$ and $\hat{u}_i(w_i^a(x_1))q_1 p_1 + \hat{u}_i(w_i^a(x_2))q_2 p_2 = 1$, respectively. By solving those two equations, we have $\hat{u}_i(w_i^a(x_1)) = -\frac{1}{q_2}$ and $\hat{u}_i(w_i^a(x_2)) = -\frac{1}{q_1}$, from which one can see that $\hat{u}_1(w_1^a(x_j)) = \hat{u}_2(w_2^a(x_j))$ for all $j = 1, 2$. Note that, as shown in the proof of Proposition 2, the two conditions that $\hat{u}_1'(0) = \hat{u}_2'(0)$ and that $-\frac{\hat{u}_1''(s)}{\hat{u}_1'(s)} \leq -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}$ for all s imply that $\hat{u}_1(s) \geq \hat{u}_2(s)$ for all s. This inequality, combined with the fact that $\hat{u}_1(w_1^a(x_j)) = \hat{u}_2(w_2^a(x_j))$ for all $j = 1, 2$, makes $\hat{u}_1(w_2^a(x_j)) \geq \hat{u}_2(w_2^a(x_j)) = \hat{u}_1(w_1^a(x_j))$ for all j, from which we have $w_1^a(x_j) \leq w_2^a(x_j)$ for all j. Therefore, one can easily see that $E[w_1^a(x)] \leq E[w_2^a(x)]$.

The Pratt's intuition in asset market that the more risk averse consumer claims the higher risk premium helps us to understand the result of Proposition 2 intuitively. Note that agent i is compensated with "risk-free" wage s_i^f in the full information case where there is no need for incentive provision, whereas he is compensated with "risky" wage $w_i^a(x) + s_i^f$ in the asymmetric information case where there exists the moral hazard problem. However, note that the principal should guarantee to him the same utility level in both cases (i.e., $E[u_i(w_i^a(x) + s_i^f)] = u_i(s_i^f)$) in order to make him participate. At this time, when risk-averse agent i is compensated with risky wage $w_i^a(x) + s_i^f$, he will demand the risk premium as much as $E[w_i^a(x)]$. This risk premium becomes the incentive cost on the principal's side. Thus, the more risk averse the agent is, the higher is the principal's incentive cost. However, it is worth noticing that this intuition holds only when the agent's incentive sensitivity is fixed.

By combining Propositions 1 and 2, we obtain the following proposition.

Proposition 3: *If i)* $\frac{u_1'(s_1^f)}{c_1'(a)} \geq \frac{u_2'(s_2^f)}{c_2'(a)}$*, and ii)* $-\frac{u_1''(s+s_1^f)}{u_1'(s+s_2^f)} \leq -\frac{u_2''(s+s_2^f)}{u_2'(s+s_2^f)}$ *for all* s*, then* $E[w_1^a(x)] \leq E[w_2^a(x)]$.

This proposition shows that the increase in an agent's incentive sensitivity and the decrease in his degree of absolute risk aversion around his fixed salary reduce the principal's incentive cost.

Proposition 3 can be easily verified by introducing an arbitrary agent k with utility $\hat{u}_k(s) = \alpha \hat{u}_2(s)$, where $\alpha = \frac{\hat{u}_1'(0)}{\hat{u}_2'(0)}$. Note that, by definition, agent k's marginal utility at 0 is equal to agent 1's (i.e., $\hat{u}_k'(0) = \hat{u}_1'(0)$). Let $w_k^a(x)$ denote his optimal incentive wage contract. Since $\hat{u}_k(s)$ is a linear transformation of $\hat{u}_2(s)$, agent k's degree of absolute risk aversion is identical to agent 2's (i.e., $-\frac{\hat{u}_k''(s)}{\hat{u}_k'(s)} = -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}$ for all s), and by condition i) in Proposition 3, agent k's marginal utility is not less than agent 2's (i.e., $\hat{u}_k'(0) \geq \hat{u}_2'(0)$). Thus, by Proposition 1, we have $E[w_k^a(x)] \leq E[w_2^a(x)]$. Furthermore, agent k's marginal utility at 0 is equal to agent 1's (i.e., $\hat{u}_k'(0) = \hat{u}_1'(0)$) and agent k's degree of absolute risk aversion is not less than agent 2's (i.e, $-\frac{\hat{u}_1''(s)}{\hat{u}_k'(s)} = -\frac{\hat{u}_k''(s)}{\hat{u}_k'(s)}$ for all s). Thus, by Proposition 2, we have $E[w_1^a(x)] \leq E[w_k^a(x)]$. Therefore, we obtain $E[w_1^a(x)] \leq E[w_2^a(x)]$.

It is important to compare our result in the principal-agent problem with the well-known Pratt's result in the asset market in order to understand ours intuitively. As explained earlier, note that the Pratt's result is that, as an consumer becomes more risk averse under the degree of absolute risk aversion, he will demand the higher risk premium. It is worth noticing that consumer's marginal utility itself does not affect his risk premium at all. On the other hand, assuming that two agents have a homogeneous cost function (i.e., $c_1(a) = c_2(a)$), our result is that, not only as the agent becomes more risk averse under the degree of absolute risk aversion, but also as the magnitude of his marginal utility decreases, his incentive cost increases. The critical difference between our result and the Pratt's result is that the participator's marginal utility itself is an important factor in our result, but not at all in the Pratt's result. This is because there is the only issue of consumer's market participation in the asset market, while, in the principal-agent problem, there are the two issues: the agent's contract participation issue and incentive provision issue. From this reason, one can see that the effect of the participator's risk aversion comes from an participation issue but the effect of his marginal utility comes from an incentive provision issue. Therefore, in the agency problem that the principal must consider the agent's participation and incentive constraints, not only the agent's risk aversion but also his marginal utility itself is important.

We have showed through Propositions 1 to 3 that the decrease in the agent's incentive sensitivity or the increase in his risk aversion around his fixed salary leads to the increase in the principal's incentive cost. This result means that, if the two agents have the same fixed salaries (i.e., $s_1^f = s_2^f$), the principal prefers the agent who is the less risk averse and/or has the higher incentive sensitivity. However, there is an interesting question: Is it possible that the principal prefers the more risk averse agent? The following proposition shows that it is possible if the more risk averse agent has the higher incentive sensitivity and the effect of incentive sensitivity dominates that of risk aversion.

Proposition 4: *Consider the two-outcome case with* $x \in \{x_1, x_2\}$ *where* $q_j = \frac{f_a(x_j|a)}{f(x_j|a)}$ *for* $j = 1, 2$. *Suppose that* $|q_1|$ *and* $|q_2|$ *are sufficiently large. Although* $-\frac{u_1''(s+s_1^f)}{u_1'(s+s_1^f)} \geq -\frac{u_1''(s+s_2^f)}{u_1'(s+s_2^f)}$ *for all* s, *if* $-\frac{u_1''(s_1^f)[c_1'(a)]^2}{[u_1'(s_1^f)]^3} < -\frac{u_2''(s_2^f)[c_2'(a)]^2}{[u_2'(s_2^f)]^3}$, *then it is possible that* $E[w_1^a(x)] \leq E[w_2^a(x)]$.

Return to the example of two-outcome case suggested for Proposition 2. Note that, since $\hat{u}_i(w_i^a(x_1)) = -\frac{1}{q_2}$ and $\hat{u}_i(w_i^a(x_2)) = -\frac{1}{q_1}$ for all i, the difference $|\hat{u}_i(w_i^a(x_1)) - \hat{u}_i(w_i^a(x_2))|$ is equal to the difference $|\frac{1}{q_1} - \frac{1}{q_2}|$. Then, if both $|q_1|$ and $|q_2|$ are sufficiently large, the difference $|\hat{u}_i(w_i^a(x_1)) - \hat{u}_i(w_i^a(x_2))|$ is sufficiently small, and so the difference $|w_i^a(x_1) - w_i^a(x_2)|$ is also sufficiently small. In other words, when $|q_1|$ and $|q_2|$ are sufficiently large, it is possible that, for all i, incentive wages $w_i^a(x_1)$ and $w_i^a(x_2)$ are optimally designed near zero. In this case, what Proposition 4 indicates is that, although agent 1 is more risk averse than agent 2 around their own fixed salaries (i.e., $-\frac{u_1''(s+s_1^f)}{u_1'(s+s_1^f)} \geq -\frac{u_1''(s+s_2^f)}{u_1'(s+s_2^f)}$ for all s), if, compared with incentive sensitivity of agent 2, that of agent 1 is larger enough to satisfy $-\frac{u_1''(s_1^f)[c_1'(a)]^2}{[u_1'(s_1^f)]^3} < -\frac{u_2''(s_2^f)[c_2'(a)]^2}{[u_2'(s_2^f)]^3}$, it is possible that agent 1 has the lower incentive cost than agent 2.

Proposition 4 has an implication that, when the principal can make sufficiently precise inference about the agent's action choice, it is possible that the principal prefer the agent with the higher incentive sensitivity regardless of his risk aversion. Note that, since $\frac{f_a(x|a)}{f(x|a)} \approx \frac{f(x|a)-f(x|a')}{f(x|a)(a-a')}$ where a' is in the neighbourhood of a, the condition that $|q_j| = |\frac{f_a(x_j|a)}{f(x_j|a)}|$, $j = 1, 2$, are sufficiently large means that, when agent's action changes from a to a', the variation of probability $|f(x|a) - f(x|a')|$ is sufficiently large, and so the principal can infer the agent's action choice very well. So, the principal can design the incentive wage contract with low power, that is, $|w_i^a(x_2) - w_i^a(x_1)|$ very small. In this case, what Proposition 4 implies is that what is more important to the principal may be the agent's incentive sensitivity rather than his risk aversion.

4 Conclusion

We have analyzed not only the relationship between the agent's incentive sensitivity and the principal's incentive cost but also the relationship between his degree of risk aversion and her incentive cost. We first showed that, when one agent's incentive sensitivity measured as the marginal rate of substitution between income and effort is higher than the other agent's, the incentive cost from the former agent is less than that from the latter agent, only when they have the same degree of absolute risk aversion around their own fixed salaries. This means that the agent's incentive sensitivity negatively affects the principal's incentive cost. Thus, in order to analyze the pure effect of the agent's risk aversion on the principal's incentive cost, it is needed that his incentive sensitivity should be fixed. Under such condition, we showed that, when the agent becomes more risk-averse under the degree of absolute risk aversion around his fixed salary, the principal's incentive cost increases. Furthermore, we have showed that, when the agent becomes more risk-averse and at the same time when his incentive sensitivity become lowered, the principal's incentive cost increases. Finally, we provided an example that the effect of the agent's incentive sensitivity on the principal's incentive cost dominates that of his risk aversion.

The effects of the agent's incentive sensitivity and his risk aversion will depend on the characteristics of output's distribution. However, we believe that it is possible to compare both effects mathematically when some specific distributions are given. Our guess is that, as the principal makes more precise inference about the agent's action choice, the effect of incentive sensitivity would grow greater but the effect of risk aversion would grow lower, and in the end, the former would dominate the latter, as in Proposition 4.

5 Appendix

Proof of Lemma 1. Recall that $\hat{u}_i(s) \equiv \frac{1}{c_i'(a)}[u_i(s+s_i^f) - u_i(s_i^f)]$ for all i. Using this utility function, the participation and the doubly relaxed incentive constraints for agent i become

$$\int \hat{u}_i(w_i(x))f(x|a)dx \geq 0, \tag{A.1}$$

and

$$\int \hat{u}_i(w_i(x))f_a(x|a)dx \geq 1. \tag{A.2}$$

Note that, for all i, (A.1) and (A.2) should be binding at the optimum.

We introduce an arbitrary contract $r(x)$ such that $\hat{u}_1(r(x)) \equiv \hat{u}_1(w_2^a(x)) - E[\hat{u}_1(w_2^a(x))]$. Since $E[\hat{u}_1(w_2^a(x))] \geq 0$ by condition ii), it is true that $r(x) \leq w_2^a(x)$ for all $x \in X$, which implies that

$$E[r(x)] \leq E[w_2^a(x)]. \tag{A.3}$$

We will prove that $E[w_1^a(x)] \leq E[r(x)]$ by showing that $r(x)$ belongs to the principal's opportunity set for $w_1(x)$ satisfying the participation and (doubly relaxed) incentive constraints for agent 1, which, together with (A.3), implies that $E[w_1^a(x)] \leq E[w_2^a(x)]$.

When $r(x)$ is offered to agent 1 with parallel-shifted utility function $\hat{u}_1(s)$, since

$$E[\hat{u}_1(r(x))] = E[\hat{u}_1(w_2^a(x))] - E[\hat{u}_1(w_2^a(x))] = 0, \tag{A.4}$$

$r(x)$ satisfies the participation constraint (A.1) for agent 1. And, combining (2) and (A.4) yields

$$E[\hat{u}_1(r(x))] = E[\hat{u}_2(w_2^a(x))],$$

which implies that $\hat{u}_1(r(x)) \equiv \hat{u}_1(w_2^a(x)) - E[\hat{u}_1(w_2^a(x))]$ must cross $\hat{u}_2(w_2^a(x))$ at once from below, since $\hat{u}_1'(s) \geq \hat{u}_2'(s)$, $\forall s$, by condition i) and since $w_2^a(x)$ is an increasing function by the MLRP for $f(x|a)$.[14] Then, since, by Lemma 1 in Innes (1990),

$$\int \hat{u}_1(r(x)) f_a(x|a) dx = \int \{\hat{u}_1(w_2^a(x)) - E[\hat{u}_1(w_2^a(x))]\} f_a(x|a) dx$$
$$\geq \int \hat{u}_2(w_2^a(x)) f_a(x|a) dx = 1,$$

where the last equality holds by (3), $r(x)$ satisfies the doubly relaxed incentive constraint (A.2) for agent 1. Thus, since $r(x)$ satisfies the participation and incentive constraints for agent 1, that is, $r(x)$ belongs to the principal's opportunity set for $w_1(x)$, and since $w_1^a(x)$ is the optimal solution of cost minimization problem for agent 1, we have

$$E[w_1^a(x)] \leq E[r(x)]. \tag{A.5}$$

Therefore, combing (A.3) and (A.5) gives $E[w_1^a(x)] \leq E[w_2^a(x)]$. □

Proof of Proposition 1. Since $\hat{u}_i(s) \equiv \frac{1}{c_i'(a)}[u_i(s + s_i^f) - u_i(s_i^f)]$, for all $i = 1, 2$, conditions i) and ii) in Proposition 1 are rewritten by

$$\hat{u}_1'(0) \geq \hat{u}_2'(0), \tag{A.6}$$

and

$$-\frac{\hat{u}_1''(s)}{\hat{u}_1'(s)} = -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}, \quad \forall s, \tag{A.7}$$

respectively.

Taking an integral on both sides of (A.7) gives

$$\ln \hat{u}_1'(s) - \ln \hat{u}_1'(0) = \ln \hat{u}_2'(s) - \ln \hat{u}_2'(0) \quad \Leftrightarrow \quad \frac{\hat{u}_1'(s)}{\hat{u}_2'(s)} = \frac{\hat{u}_1'(0)}{\hat{u}_2'(0)} \equiv \alpha,$$

where $\alpha \geq 1$ by (A.6), which is equal to

$$\hat{u}_1'(s) = \alpha \hat{u}_2'(s), \quad \forall s,$$

[14] As seen in (1), the MLRP that $\frac{f_a(x|a)}{f(x|a)}$ should be increasing in x for all a implies that $w_i^a(x)$ is increasing in x.

of which integrating both sides gives

$$\hat{u}_1(s) - \hat{u}_1(0) = \alpha[\hat{u}_2(s) - \hat{u}_2(0)].$$

Then, since $\hat{u}_i(0) = \frac{1}{c_i'(a)}[u_i(s_i^f) - u_i(s_i^f)] = 0$ for all $i = 1, 2$, we have

$$\hat{u}_1(s) = \alpha \hat{u}_2(s).$$

Note that $E[\hat{u}_1(w_2^a(x))] = \alpha E[\hat{u}_2(w_2^a(x))] = 0$, where the second equality holds by (2), and that, since $\alpha \equiv \frac{\hat{u}_1'(0)}{\hat{u}_2'(0)} \geq 1$, $\hat{u}'(s) \geq \hat{u}_2'(s)$ for all s. Therefore, by Lemma 1, we have $E[w_1^a(x)] \leq E[w_2^a(x)]$. $\square$

Proof of Proposition 2. Note that, by using $\hat{u}_i(s) \equiv \frac{1}{c_i'(a)}[u_i(s + s_i^f) - u_i(s_i^f)]$, conditions i) and ii) in Proposition 2 are rewritten by

$$\hat{u}_1'(0) = \hat{u}_2'(0), \tag{A8}$$

and

$$-\frac{\hat{u}_1''(s)}{\hat{u}_1'(s)} \leq -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}, \quad \forall s, \tag{A9}$$

respectively.

Firstly, we will show that (A8) and (A9) implies that $\hat{u}_1(s) \geq \hat{u}_2(s)$ for all s. For any $s < 0$, integrating both sides of (A9) from s to 0 gives

$$\ln \frac{\hat{u}_1'(s)}{\hat{u}_1'(0)} \leq \ln \frac{\hat{u}_2'(s)}{\hat{u}_2'(0)} \quad \Leftrightarrow \quad \hat{u}_1'(s) \leq \hat{u}_2'(s), \quad \forall s < 0, \tag{A10}$$

where the equivalence holds by (A8). And, integrating both sides of (A10) from s to 0 gives

$$\hat{u}_1(0) - \hat{u}_1(s) \leq \hat{u}_2(0) - \hat{u}_2(s),$$

which, together with the fact that $\hat{u}_1(0) = \hat{u}_2(0) = 0$, implies that

$$\hat{u}_1(s) \geq \hat{u}_2(s), \quad \forall s < 0, \tag{A11}$$

Similarly, for any $s \geq 0$, integrating both sides of (A9) from 0 to s gives

$$\ln \frac{\hat{u}_1'(0)}{\hat{u}_1'(s)} \leq \ln \frac{\hat{u}_2'(0)}{\hat{u}_2'(s)} \quad \Leftrightarrow \quad \hat{u}_1'(s) \geq \hat{u}_2'(s), \quad \forall s \geq 0, \tag{A12}$$

where the equivalence holds by (A8). Again, integrating both sides of (A12) from 0 to s gives

$$\hat{u}_1(s) - \hat{u}_1(0) \geq \hat{u}_2(s) - \hat{u}_2(0),$$

which, by (A8), implies that

$$\hat{u}_1(s) \geq \hat{u}_2(s), \quad \forall s \geq 0. \tag{A13}$$

Thus, by (A11) and (A13), we have

$$\hat{u}_1(s) \geq \hat{u}_2(s), \quad \forall s. \tag{A14}$$

Define $\nu_i(z) \equiv argmax_s s - z\hat{u}_i(s)$ and $\psi_i(z) \equiv \nu_i(z) - z\hat{u}_i(\nu_i(z))$ for any given $z > 0$. Note that applying Lemma 4 of Jewitt, Kadan and Swinkels (2008) to the cost minimization problem for agent i with parallel-shifted utility $\hat{u}_i(s)$ gives

$$\begin{aligned}
E[w_i^a(x)] &= \max_{\lambda,\mu\geq 0} \min_s \int \Big\{ s - \big(\lambda + \mu \frac{f_a(x|a)}{f(x|a)}\big) \times \hat{u}_i(s) \Big\} f(x|a)dx + \mu \\
&= \max_{\lambda,\mu\geq 0} L_i(\lambda,\mu;a) \equiv \int \psi_i\Big(\lambda + \mu \frac{f_a(x|a)}{f(x|a)}\Big) f(x|a)dx + \mu \\
&= L_i(\hat{\lambda}_i, \hat{\mu}_i; a),
\end{aligned}$$

where the second equality comes from the definition of $\psi(z)$. Since $L_2(\hat{\lambda}_2, \hat{\mu}_2; a) = max_{\lambda,\mu\geq 0} L_2(\lambda,\mu;a)$, we have

$$L_2(\hat{\lambda}_2, \hat{\mu}_2; a) \geq L_2(\hat{\lambda}_1, \hat{\mu}_1; a).$$

Then,

$$\begin{aligned}
E[w_2^a(x)] - E[w_1^a(x)] &= L_2(\hat{\lambda}_2, \hat{\mu}_2; a) - L_1(\hat{\lambda}_1, \hat{\mu}_1; a) \\
&\geq L_2(\hat{\lambda}_1, \hat{\mu}_1; a) - L_1(\hat{\lambda}_1, \hat{\mu}_1; a) \\
&= \int \Big[\psi_2\big(\hat{\lambda}_1 + \hat{\mu}_1 \frac{f_a(x|a)}{f(x|a)}\big) - \psi_1\big(\hat{\lambda}_1 + \hat{\mu}_1 \frac{f_a(x|a)}{f(x|a)}\big) \Big] f(x|a)dx. \qquad \text{(A15)}
\end{aligned}$$

For any $z > 0$, since $\nu_1(z) = argmax_s s - z\hat{u}_1(s)$ and $\psi_1(z) \equiv \nu_1(z) - z\hat{u}_1(\nu_1(z))$, we have

$$\psi_1(z) = \nu_1(z) - z\hat{u}_1(\nu_1(z)) \leq \nu_2(z) - z\hat{u}_1(\nu_2(z)).$$

Thus,

$$\begin{aligned}\psi_2(z) - \psi_1(z) &= \nu_2(z) - z\hat{u}_2(\nu_2(z)) - [\nu_1(z) - z\hat{u}_1(\nu_1(z))] \\ &\geq \nu_2(z) - z\hat{u}_2(\nu_2(z)) - [\nu_2(z) - z\hat{u}_1(\nu_2(z))] \\ &= z[\hat{u}_1(\nu_2(z)) - \hat{u}_2(\nu_2(z))] \geq 0, \quad \forall z > 0, \end{aligned} \tag{A16}$$

where the last inequality holds from the facts that $z > 0$ and that $\hat{u}_1(s) \geq \hat{u}_2(s)$ for all s from (A14).

Since $\hat{\lambda}_1 + \hat{\mu}_1 \frac{f_a(x|a)}{f(x|a)} > 0$ for almost every x from the existence of the optimal contract $w_1^a(x)$, applying (A16) to (A15) gives $E[w_2^a(x)] - E[w_1^a(x)] \geq 0$. $\square$

Proof of Proposition 3. We introduce a new agent k with utility function $\hat{u}_k(s) = \alpha\hat{u}_2(s)$, where $\alpha \equiv \frac{\hat{u}_1'(0)}{\hat{u}_2'(0)} = \frac{u_1'(s_1^f)c_2'(a)}{u_2'(s_2^f)c_1'(a)} \geq 1$ by condition i) in Proposition 3. Note that $\hat{u}_k(0) = \hat{u}_2(0) = 0$ by (4).

Since $\hat{u}_k'(0) = \alpha\hat{u}_2'(0) \geq \hat{u}_2'(0)$ from the fact that $\alpha \geq 1$, and since

$$-\frac{\hat{u}_k''(s)}{\hat{u}_k'(s)} = -\frac{\hat{u}_2''(s)}{\hat{u}_2'(s)}, \quad \forall s, \tag{A.17}$$

by Proposition 2, we have

$$E[w_k^a(x)] \leq E[w_2^a(x)], \tag{A.18}$$

where $w_k^a(x)$ denotes the optimal contract derived solving cost minimization problem for agent k. On the other hand, since combining (A.17) and condition ii) in Proposition 3 yields

$$-\frac{\hat{u}_1''(s)}{\hat{u}_1'(s)} \leq -\frac{\hat{u}_k''(s)}{\hat{u}_k'(s)},$$

and since $\hat{u}_k'(0) = \alpha\hat{u}_2'(0) = \hat{u}_1'(0)$, by Proposition 2 we have

$$E[w_1^a(x)] \leq E[w_k^a(x)]. \tag{A.19}$$

Therefore, by (A.18) and (A.19) we have $E[w_1^a(x)] \leq E[w_2^a(x)]$. $\square$

Proof of Proposition 4. Consider a distribution $f(x|a)$ with outcome $x \in X = \{x_1, x_2\}$. For $j = 1, 2$, let $p_j = f(x_j|a)$ and $q_j = \frac{f_a(x_j|a)}{f(x_j|a)}$. Note that $p_1 + p_2 = 1$ since $\sum_j f(x_j|a) = 1$ and note that $q_1 p_1 + q_2 p_2 = 0$ since $\sum_j \frac{f_a(x_j|a)}{f(x_j|a)} f(x_j|a) = \sum_j f_a(x_j|a) = 0$

For every agent $i = 1, 2$, since his participation constraint and incentive constraint should be binding at the optimum, we have

$$u_i(w_i^a(x_1) + s_i^f)p_1 + u_i(w_i^a(x_2) + s_i^f)p_2 = u_i(s_i^f), \tag{A.26}$$

and

$$u_i(w_i^a(x_1) + s_i^f)q_1 p_1 + u_i(w_i^a(x_2) + s_i^f)q_2 p_2 = c_i'(a), \tag{A.27}$$

respectively. Solving (A.26) and (A.27) gives

$$\begin{aligned} & u_i(w_i^a(x_1) + s_i^f) = u_i(s_i^f) - \frac{c_i'(a)}{q_2} \\ \Leftrightarrow \quad & w_i^a(x_1) = u_i^{-1}\Big(u_i(s_i^f) - \frac{c_i'(a)}{q_2}\Big) - s_i^f, \end{aligned} \tag{A.28}$$

and

$$\begin{aligned} & u_i(w_i^a(x_2) + s_i^f) = u_i(s_i^f) - \frac{c_i'(a)}{q_1} \\ \Leftrightarrow \quad & w_i^a(x_2) = u_i^{-1}\Big(u_i(s_i^f) - \frac{c_i'(a)}{q_1}\Big) - s_i^f. \end{aligned} \tag{A.29}$$

Let $\eta_i(x) = u_i^{-1}(u_i(s_i^f) + x)$ for $i = 1, 2$. Note that $\eta_i(0) = u_i^{-1}(u_i(s_i^f)) = s_i^f$, $\eta_i'(0) = \frac{1}{u_i'(s_i^f)}$, and $\eta_i''(0) = -\frac{u_i''(s_i^f)}{[u_i'(s_i^f)]^3}$. Since, if $|x|$ is sufficiently small, $\eta_i(x)$ is approximated by $\eta_i(x) = \eta_i(0) + \eta_i'(0)x + \frac{\eta_i''(0)}{2}x^2$, that is,

$$u_i^{-1}(u_i(s_i^f) + x) \approx s_i^f + \frac{1}{u_i'(s_i^f)}x - \frac{u_i''(s_i^f)}{2[u_i'(s_i^f)]^3}x^2,$$

(A.28) and (A.29) are transformed into

$$w_i^a(x_1) = \eta_i\Big(\frac{c_i'(a)}{q_2}\Big) - s_i^f \approx \frac{1}{u_i'(s_i^f)} \times \frac{c_i'(a)}{q_2} - \frac{u_i''(s_i^f)}{2[u_i'(s_i^f)]^3} \times \frac{[c_i'(a)]^2}{q_2^2},$$

and

$$w_i^a(x_2) = \eta_i\Big(\frac{c_i'(a)}{q_1}\Big) - s_i^f \approx \frac{1}{u_i'(s_i^f)} \times \frac{c_i'(a)}{q_1} - \frac{u_i''(s_i^f)}{2[u_i'(s_i^f)]^3} \times \frac{[c_i'(a)]^2}{q_1^2}.$$

Then, we have

$$E[w_i^a(x)] \approx -\frac{u_i''(s_i^f)[c_i'(a)]^2}{[u_i'(s_i^f)]^3} \times \frac{1}{2}\Big(\frac{p_1}{q_2^2} + \frac{p_2}{q_1^2}\Big) \tag{5}$$

Thus, when $|q_1|$ and $|q_2|$ are sufficiently large, if $-\frac{u_1''(s_1^f)[c_1'(a)]^2}{[u_1'(s_1^f)]^3} < -\frac{u_2''(s_2^f)[c_2'(a)]^2}{[u_2'(s_2^f)]^3}$, it is possible that $E[w_1^a(x)] \leq E[w_2^a(x)]$. □

References

[1] Chade, H., and de Serio, V. N. V. (2002): "Risk aversion, moral hazard, and the principal's loss," *Economic Theory*, 20(3), 637-644.

[2] Grossman, S. and O. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7-45.

[3] Innes, R. D. (1990): "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, 52, 45-67.

[4] Jewitt, I. (1997): "Information and Principal-Agent Problems," Working paper, University of Bristol.

[5] Jewitt, I., O. Kadan, and J. M. Swinkels (2008): "Moral Hazard with bounded payments," *Journal of Economic Theory*, 143(1), 59-82.

[6] Jung, J.Y. and S.K. Kim (2015): "Information Space Conditions for the First-Order Approach in Agency Problems," *Journal of Economic Theory*, 160, 243-279.

[7] Kim, S. K. (1995): "Efficiency of an Information System in an Agency Model," *Econometrica*, 63, 89-102.

[8] Mornoni, S. and J. Swinkels (2014): "Existence and non-existence in the moral hazard problem," *Journal of Economic Theory*, 150, 668-682.